PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/00

A1

(11) International Publication Number: WO 99/50402

(43) International Publication Date: 7 October 1999 (07.10.99)

(21) International Application Number: PCT/US99/06139

(22) International Filing Date: 26 March 1999 (26.03.99)

(30) Priority Data:
60/079,770 27 March 1998 (27.03.98) US

(71) Applicant: PRESIDENT AND FELLOWS OF HARVARD COLLEGE [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: MEKALANOS, John, J.; 78 Fresh Pond Lane, Cambridge, MA 02138 (US). AKERLEY, Brian; Apartment #2, 74 St. Paul Street, Brookline, MA 02146 (US). RUBIN, Eric; 283 Woodward Street, Waban, MA 02168 (US). CAMILLI, Andrew; 5 Moose Hill Parkway, Sharon, MA 02067 (US).

(74) Agent: BIEKER-BRADY, Kristina; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110-2214 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: SYSTEMATIC IDENTIFICATION OF ESSENTIAL GENES BY *IN VITRO* TRANSPOSON MUTAGENESIS

(57) Abstract

The invention features a general system for the identification of essential genes in organisms. This system is applicable to the discovery of novel target genes for antimicrobial compounds, as well as to the discovery of genes that enhance cell growth or viability.

SYSTEMATIC IDENTIFICATION OF ESSENTIAL GENES BY *IN VITRO* TRANSPOSON MUTAGENESIS

Statement as to Federally Sponsored Research

This research has been sponsored in part by NIH grants AI02137 and AI26289. The government has certain rights to the invention.

Background of the Invention

Nearly 40% of the *Haemophilus (H.) influenzae* genome is comprised of genes of unknown function, many of which have no recognizable functional orthologues in other species. Similar numbers of unidentified open reading frames (orfs) are present in other sequenced or partially sequenced genomes of infectious organisms. Comprehensive screens and selections for identifying functional classes of genes provide a crucial starting point for converting the vast body of growing sequence data into meaningful biological information that can be used for drug discovery.

One major and important class of genes consists of those bacterial genes that are essential for growth or viability of a bacterium. Because useful conventional antibiotics are known to act by interfering with the products of essential genes, it is likely that the discovery of new essential gene products will have a significant impact on efforts to develop novel antimicrobial drugs. Essential gene products have been traditionally identified through the isolation of conditional lethal mutants, or by transposon mutagenesis in the presence of a complementing wild type allele (balanced lethality). However, such approaches are laborious, as they require identification, purification, and study of individual mutant strains. These methods are also limited to species with

well-developed systems for genetic manipulation and, therefore, cannot be readily applied to many of the potentially dangerous microorganisms whose genomes have recently been sequenced.

In order to facilitate the discovery of novel anti-microbial drugs, it would be desirable to have a rapid, generalized method of identifying essential growth/viability genes in pathogens. Such a method would be particularly useful for identifying essential genes in pathogens that are not genetically well-characterized. Such a method could also be used to identify essential genes in higher organisms, e.g., in animals and in plants.

Summary of the Invention

We have developed a general system for the identification of essential genes in organisms. The system may be used to discover novel target genes for the development of therapeutic compounds, as well as for the discovery of genes that are involved in cell growth or viability. A related aspect of the invention allows for rapid construction of conditional mutations in essential genes.

In general, the invention features a method for locating an essential region in a portion of DNA from the genome of an organism. The method includes: a) mutagenizing DNA having the sequence of an essential portion of DNA, wherein the mutagenizing is performed using *in vitro* mutagenesis with a transposon; b) transforming cells of the organism with the mutagenized DNA of step a); c) identifying cells containing the mutagenized DNA; and d) locating the essential region of the DNA portion by detecting the absence of transposons in the essential region of DNA in cells containing the mutagenized DNA.

In various embodiments, the transposon may contain a selectable marker, the transposon may be *mariner*, and the method may further comprise

the use of *Himar 1* transposase.

In a preferred embodiment, the *in vitro* mutagenesis is high saturation mutagenesis. In further embodiments, the portion of DNA may be amplified using the polymerase chain reaction (PCR) prior to mutagenesis, or the portion of DNA may be cloned into a vector prior to mutagenesis. In another embodiment, prior to transforming the cells, the mutagenized DNA may be subjected to gap repair using DNA polymerase and DNA ligase. In still another embodiment, the transposon-mutagenized DNA may be recombined into the chromosome using an allelic replacement vector.

In another preferred embodiment, the locating of an essential region of DNA is done by performing PCR footprinting on a pool of transposon-mutagenized cells. The PCR footprinting is performed using a primer that hybridizes to the transposon, plus a primer that hybridizes to a specific location on the chromosome, after which the PCR products are separated on a footprinting gel. A PCR product on the gel represents a region of the chromosome that does not contain an essential gene, and the lack of a PCR product in an area of the gel, where a PCR product is expected, represents a region of the chromosome that contains an essential gene. Alternatively, a low level of the PCR product on the gel, relative to other PCR products on the gel, represents a region of the chromosome that contains an essential gene.

In still other embodiments, the cell may have a haploid growth phase, or be a single-cell microorganism, or be naturally competent for transformation, or be made competent for transformation, or be a fungus, such as a yeast (e.g., *Saccharomyces cerevisiae)*, or be a bacterium, including, but not limited to, a gram-positive bacterium. In a preferred embodiment, the bacterium is to be selected from the group consisting of: *Actinobacillus actinomycetemcomitans*; *Borrelia burgdorferi; Chlamydia trachomatis;*

Enterococcus faecalis; Escherichia coli; Haemophilus influenzae; Helicobacter pylori; Legionella pneumophila; *Mycobacterium avium; Mycobacterium tuberculosis*; *Mycoplasma genitalium*; *Mycoplasma pneumonia*; *Neisseria gonorrhoeae*; *Neisseria meningitidis*; *Staphylococcus aureus*; *Streptococcus pneumoniae*; *Streptococcus pyogenes*; *Treponema pallidum*; and *Vibrio cholerae*.

In another embodiment, the transposon may contain a selectable marker gene, and identifying the cells containing mutagenized DNA may be based upon the ability of the cells to grow on selective medium, wherein a cell containing a transposon can grow on selective medium, and a cell lacking a transposon cannot grow, or grows more slowly, on selective medium.

In still another embodiment, the transposon may contain a reporter gene, and identifying cells containing mutagenized DNA may be based on a reporter gene assay, wherein a cell confirming a transposon expresses the reporter genc and a cell lacking a transposon does not express the reporter gene.

In yet another embodiment, the method includes a step in which the cells are cultured in a medium that approximates a host environment for a pathogen.

In a second aspect, the invention provides a method for obtaining conditional mutations in essential genes. The method includes the steps of amplifying DNA containing a selective marker, as described herein, near an essential gene (e.g., a transposon) using mutagenic amplification (e.g., mutagenic PCR), transforming the DNA into a competent host under conditions allowing selection for those strains containing the selective marker, and screening for strains under permissive and non-permissive conditions such that conditional lethal mutations may be identified.

In a third aspect, the invention provides a method for isolating a

compound that modulates the expression of a nucleic acid sequence operably linked to a gene promoter. The method includes a) providing a cell expressing a nucleic acid sequence operably linked to a gene promoter, wherein the gene promoter is the gene promoter for: HI0455; HI0456; HI0458; HI0599; HI0887; HI0904; HI0906; HI0907; HI0908; HI0909; HI1650; HI1651; HI1654; HI1655; *S. pneumoniae* rbfA; *S. pneumoniae* IF-2; *S. pneumoniae* L7AE; or *S. pneumoniae* nusA; b) contacting the cell with a candidate compound; and c) detecting or measuring expression of the gene following contact of the cell with the candidate compound.

In preferred embodiments of the third aspect, the nucleic acid sequence is a reporter gene (e.g., GFP, lacZ, or alkaline phosphatase) or is HI0455; HI0456; HI0458; HI0599; HI0887; HI0904; HI0906; HI0907; HI0908; HI0909; HI1650; HI1651; HI1654; HI1655; *S. pneumoniae* rbfA; *S. pneumoniae* IF-2; *S. pneumoniae* L7AE; or *S. pneumoniae* nusA.

In yet another preferred embodiment of the third aspect, the modulation in the expression of the nucleic acid sequence modulates cell growth or viability of the cell.

In a fourth aspect, the invention provides a method for identifying a nucleic acid sequence that is essential for cell growth or viability. The method includes a) expressing in a cell (i) a first nucleic acid sequence operably linked to a gene promoter, wherein the gene promoter is the gene promoter for: HI0455; HI0456; HI0458; HI0599; HI0887; HI0904; HI0906; HI0907; HI0908; HI0909; HI1650; HI1651; HI1654; HI1655; *S. pneumoniae* rbfA; *S. pneumoniae* IF-2; *S. pneumoniae* L7AE; or *S. pneumoniae* nusA; and (ii) a second nucleic acid sequence; and b) monitoring the expression of the first nucleic acid sequence, wherein an increase in the expression identifies the second nucleic acid sequence as being essential for cell growth or viability.

In preferred embodiments of the fourth aspect, the first nucleic acid sequence is a reporter gene (eg., GFP, lacZ, or alkaline phosphatase), or is HI0455; HI0456; HI0458; HI0599; HI0887; HI0904; HI0906; HI0907; HI0908; HI0909; HI1650; HI1651; HI1654; HI1655; *S. pneumoniae* rbfA; *S. pneumoniae* IF-2; *S. pneumoniae* L7AE; or *S. pneumoniae* nusA.

In another embodiment of the fourth aspect, the increase in the expression of the nucleic acid sequence increases cell growth or viability of the cell.

In preferred embodiments of the third or fourth aspect, the expression nucleic acid sequence is measured by assaying the protein level or the RNA level of the nucleic acid sequence.

In other preferred embodiments of the third or fourth aspect, the cell is a single-cell microorganism or the microorganism is a bacterium (e.g., a gram-positive bacterium). A preferred bacterium is one that is selected from the group consisting of: *Actinobacillus actinomycetemcomitans*; *Borrelia burgdorferi; Chlamydia trachomatis; Enterococcus faecalis; Escherichia coli; Haemophilus influenzae; Helicobacter pylori; Legionella pneumophila*; *Mycobacterium avium; Mycobacterium tuberculosis*; *Mycoplasma genitalium*; *Mycoplasma pneumonia*; *Neisseria gonorrhoeae*; *Neisseria meningitidis*; *Staphylococcus aureus*; *Streptococcus pneumoniae*; *Streptococcus pyogenes*; *Treponema pallidum*; and *Vibrio cholerae.*

By "cells of an organism" is meant cells that undergo homologous recombination. Such cells may be of bacterial, mycobacterial, yeast, fungal, algal, plant, or animal origin.

By "homologous recombination" is meant a process by which an exogenously introduced DNA molecule integrates into a target DNA molecule in a region where there is identical or near-identical nucleotide sequence

between the two molecules. Homologous recombination is mediated by complementary base-pairing, and may result in either insertion of the exogenous DNA into the target DNA (a single cross-over event), or replacement of the target DNA by the exogenous DNA (a double cross-over event). Such events may occur in virtually any normal cell, including bacterial, mycobacterial, yeast, fungal, algal, plant, or animal cells.

By "transposon" is meant a DNA molecule that is capable of integrating into a target DNA molecule, without sharing homology with the target DNA molecule. The target molecule may be, for example, chromosomal DNA, cloned DNA, or PCR-amplified DNA. Transposon integration is catalyzed by transposase enzyme, which may be encoded by the transposon itself, or may be exogenously supplied. One example of a transposon is *mariner*. Other examples include Tn5, Tn7 and Tn10.

By "*in vitro* transposition" is meant integration of a transposon into target DNA that is not within a living cell. In an *in vitro* transposition reaction, the transposon integrates into the target DNA randomly, or with near randomness; that is, all DNA regions in the target DNA have approximately equal chances of being sites for transposon integration.

By "selectable marker" is meant a gene carried by a transposon that alters the ability of a cell harboring the transposon to grow or survive in a given growth environment relative to a similar cell lacking the selectable marker. Such a marker may be a positive or negative selectable marker. For example, a positive selectable marker (e.g., an antibiotic resistance or auxotrophic growth gene) encodes a product that confers growth or survival abilities in selective medium (e.g., containing an antibiotic or lacking an essential nutrient). A negative selectable marker, in contrast, prevents transposon-harboring cells from growing in negative selection medium, when compared to cells not

harboring the transposon. A selectable marker may confer both positive and negative selectability, depending upon the medium used to grow the cell. The use of selectable markers in prokaryotic and eukaryotic cells is well known by those of skill in the art.

By "permissive growth conditions" or "rich growth conditions" is meant an environment that is relatively favorable for cell growth and/or viability. Such conditions take into account the relative availability of nutrients, the absence of toxins, and optimal temperature, atmospheric pressure, presence or absence of gases (such as oxygen and carbon dioxide), and exposure to light, as required by the organism being studied. Permissive growth conditions may exist *in vitro* (such as in liquid and on solid culture media) or *in vivo* (such as in the natural host or environment of the cell being studied).

By "stringent growth conditions" is meant an environment that is relatively unfavorable for growth and/or viability of cells of an organism. An unfavorable environment may be due to nutrient limitations (e.g., as seen with "minimal" bacterial growth medium such as MIc), the presence of a compound that is toxic for the cell under study, an environmental temperature, gas concentration, light intensity, or atmospheric pressure that is extreme (e.g., either too high or too low) for optimal growth/viability of the organism under study.

By "gene that is essential for growth and/or viability" or by "essential gene" or by "essential region in a portion of DNA" is meant a DNA element such as an origin of replication or a gene that encodes a polypeptide or RNA whose function is required for survival, growth, or mitosis/meiosis of a cell. Insertion of a transposon into an essential gene may be lethal, i.e., prevent a cell from surviving, or it may prevent a cell from growing or undergoing

mitosis/meiosis. Alternatively, insertion of a transposon into an essential gene may allow survival of a cell, but result in severely diminished growth or metabolic rate. An essential gene also may be conditionally essential (i.e., required for viability and/or growth under certain conditions, but not under other conditions).

By "absence of transposons" is meant that fewer transposon insertions are detected in an essential region of DNA, relative to the number of transposon insertions detected in a non-essential region of DNA. An absence of transposons may be absolute (i.e., zero transposons detected) or relative (i.e., fewer transposons detected).

By "transformation" is meant any method for introducing foreign molecules, such as DNA, into a cell. Lipofection, DEAE-dextran-mediated transfection, microinjection, protoplast fusion, calcium phosphate precipitation, retroviral delivery, electroporation, natural transformation, and biolistic transformation are just a few of the methods known to those skilled in the art which may be used. For example, biolistic transformation is a method for introducing foreign molecules into a cell using velocity driven microprojectiles such as tungsten or gold particles. Such velocity-driven methods originate from pressure bursts which include, but are not limited to, helium-driven, air-driven, and gunpowder-driven techniques. Biolistic transformation may be applied to the transformation or transfection of a wide variety of cell types and intact tissues including, without limitation, intracellular organelles (e.g., and mitochondria and chloroplasts), bacteria, yeast, fungi, algae, plant tissue, cultured cells, and animal tissue and cultured cells.

By "identifying cells containing mutagenized DNA" is meant exposing the population of cells transformed with transposon-mutagenized DNA to selective pressure (such as growth in the presence of an antibiotic or

the absence of a nutrient) consistent with a selectable marker carried by the transposon (e.g., an antibiotic resistance gene or auxotrophic growth gene known to those skilled in the art). Identifying cells containing mutagenized DNA may also be done by subjecting transformed cells to a reporter gene assay for a reporter gene product encoded by the transposon. Selections and screens may be employed to identify cells containing mutagenized DNA, although selections are preferred.

By "reporter gene" is meant any gene which encodes a product whose expression is detectable and/or quantitatable by immunological, chemical, biochemical, biological, or mechanical assays. A reporter gene product may, for example, have one of the following attributes, without restriction: fluorescence (e.g., green fluorescent protein), enzymatic activity (e.g., lacZ/β-galactosidase, luciferase, chloramphenicol acetyltransferase, alkaline phosphatase), toxicity (e.g., ricin), or an ability to be specifically bound by a second molecule (e.g., biotin or a detectably labelled antibody). It is understood that any engineered variants of reporter genes, which are readily available to one skilled in the art, are also included, without restriction, in the foregoing definition.

By "allelic replacement vector" is meant any DNA element that can be used to introduce mutations into the genome of a target cell by specific replacement of a native gene with a mutated copy. For example, gene replacement in bacteria is commonly performed using plasmids that contain a target gene containing a mutation and a negative selectable marker outside of the region of homology. Such a plasmid integrates into the target chromosome by homologous recombination (single cross-over). Appropriate selection yields cells that have lost the negative selection marker by a second homologous recombination event (double cross-over) and contain only a mutant copy of the

target gene.

By "high saturation mutagenesis" is meant a transposon insertion frequency of at least three insertions per kilobase of target DNA, preferably, at least four insertions per kilobase of target DNA, more preferably at least five or six insertions per kilobase, and most preferably, at least seven or eight transposon insertions per kilobase of target DNA.

By "locating an essential region in a portion of DNA" is meant determining that a given stretch of DNA contains a gene that is necessary for cell growth and/or viability. Such a gene may be necessary under all, or only under some (e.g., stringent) growth conditions. The locating may be done, for example, by PCR footprinting.

The invention provides a method for the rapid identification of essential or conditionally essential DNA segments. The method is applicable to any species of cell (e.g., microbial, fungal, algal, plant, animal) that is capable of being transformed by artificial means, for example, by electroporation, liposomes, calcium phosphate, DEAE dextran, calcium chloride, etc., and is capable of undergoing homologous DNA recombination. This system offers an enhanced means of ascribing important functions to the growing number of uncharacterized genes catalogued in sequence databases.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

Brief Description of the Drawings

Fig. 1A shows the strategy for producing chromosomal mutations using *in vitro* transposition mutagenesis.

Fig. 1B shows a Southern blot analysis of *H. influenzae* transposon

mutants. Genomic DNA was isolated from 16 individual mutants and was digested with *Ase*I, which cleaves once within *magellan1*. Digested DNA was subjected to agarose gel electrophoresis, transferred to nitrocellulose, and then hybridized with a probe composed solely of *magellan1* minitransposon-derived DNA.

Fig. 2 shows a schematic diagram of PCR footprinting for detection of essential genes. Target DNA mutagenized *in vitro* with the *Himar1* transposon was introduced into bacteria by transformation and homologous recombination. Recombinants were selected for drug resistance encoded by the transposon, and insertions in essential genes were lost from the pool during growth. PCR with primers that hybridized to the transposon and to specific chromosomal sites yielded a product corresponding to each mutation in the pool. DNA regions containing no insertions yielded a blank region on electrophoresis gels.

Figs. 3A-3G show genetic footprinting of *H. influenzae* mutant pools. Genetic footprinting was carried out by using a *Himar1*-specific primer and a chromosomal primer. In Fig. 3A, the positions of molecular weight standards are indicated; other panels are labeled with locus names by HI number. In Fig. 3C and 3D, cells were selected on BXV, MIc, or BXV containing trimethoprim ("Tri"). In Fig 3F, *in vitro* mutagenesis of a chromosomal fragment that included the secA gene was performed, and the mutagenized DNA was transformed into both wild-type *H. influenzae* and an *H. influenzae* strain containing pSecA.

Fig. 4 shows *H. influenzae* orfs analyzed using *in vitro* transposition mutagenesis. Orfs with essential functions are shown in black, orfs that are non-essential are shown in white, and orfs in which mutations produce growth attenuation are shown in gray. The direction of transcription for each orf is

shown along with the TIGR designation below the orf and the closest homologue above the orf. The * designates essential orfs which can sustain a very limited number of discrete insertions (<2/kbp). Conserved hypothetical orfs of unknown function are designated CH.

Figs. 5A-5R show the nucleotide and polypeptide sequence of genes found using in vitro transposition mutagenesis to be essential genes.

Fig. 6 shows a diagram depicting the identification of a gene that is essential for growth under stringent versus permissive growth conditions.

Detailed Description of the Invention

Here we describe a simple system for performing transposon mutagenesis to rapidly identify essential or conditionally essential DNA segments. The technique, termed GAMBIT (Genomic Analysis and Mapping By *in vitro* Transposition), combines extended-length PCR, *in vitro* transposition, and PCR footprinting, to screen for genes required for growth. This system takes advantage of the ability of naturally competent cells such as bacteria to efficiently take up DNA added to cultures and incorporate it by homologous recombination into their chromosome. Since mutagenesis is conducted *in vitro*, there are no host-specific steps in the procedure, making it generally applicable to any naturally transformable species.

The first step in the development of the GAMBIT method was to develop an *in vitro* mutagenesis protocol that could be used on isolated chromosomal DNA derived from a naturally competent bacterial species (Fig. 1A). To test our system we chose *H. influenzae* and *Streptococcus (S.) pneumoniae*, both of which are transformable, as test organisms, and the *mariner* transposon *Himar1*, originally isolated from the horn fly, *Haemotobia irritans* (D.J. Lampe et al., *EMBO J.* 15:5470-5479 (1996); herein incorporated

by reference). As will be described in detail below, GAMBIT analysis of ~50 kilobases of *H. influenzae* and 10 kilobases of *S. pneumoniae* DNA confirmed the essential nature of nine of nine known essential genes.

The *mariner* transposon offers two advantages. First, *mariner* transposition occurs efficiently *in vitro* and does not require cellular cofactors. Second, under the conditions we used, *mariner* shows very little insertion site specificity, requiring only the dinucleotide TA in the target sequence (and even this minor site specificity can be easily altered using different *in vitro* reaction conditions).

Chromosomal DNA was isolated and mutagenized with the *Himar1* transposase and an artificial minitransposon encoding the gene for either kanamycin (*magellan1*) or chloramphenicol (*magellan2*) resistance. Insertion of the transposon produces a short single-stranded gap on either end of the insertion site. Since *H. influenzae* and *S. pneumoniae* are known to take up single stranded DNA, these gaps required repair (using a DNA polymerase and a DNA ligase) to produce the flanking DNA sequence required for recombination into the chromosome. The mutagenized DNA was transformed into bacteria, and cells which had acquired transposon insertions by homologous recombination were selected on the appropriate antibiotic-containing medium.

Using this method, we were able to produce libraries with ~ 9,000 *H. influenzae* mutants and ~100,000 *S. pneumoniae* mutants, indicating, as predicted, that this approach is equally effective in gram-positive and gram-negative bacteria. Southern blot analysis of *Ase*I-digested DNA from 16 individual *H. influenzae* transposon mutants (Fig. 1B) revealed that each had only a single transposon insertion and that the transposon could insert at a variety of sites. Mutagenesis of *H. influenzae* using *in vitro* transposition has

been recently described using Tn7, although it has not previously been applied to gram-positive organisms.

Although mutant libraries such as those created by the above steps are quite useful for obtaining a given mutant, the GAMBIT technique works best with a greater degree of saturation of mutations to yield a high-density insertion map of a given chromosomal region. To conduct such highly-saturated mutagenesis we targeted specific genomic segments for transposition. First, oligonucleotide primers were synthesized and used to amplify ~10 kb regions of the chromosome, using the polymerase chain reaction (PCR). The resulting PCR products were purified and used as templates for *in vitro mariner* transposon mutagenesis. Each mutagenized pool of DNA was transformed into competent bacteria and plated on rich medium containing appropriate antibiotic, resulting in libraries of ~400-800 mutants, all of which contained insertions within the target chromosomal segment.

The position of each of these insertion mutations with respect to any given PCR primer, designed from genome sequence data, can then be assessed by PCR footprinting (or similar procedures) conducted on the entire pool of mutants, using a primer which hybridizes to the transposon and another primer which hybridizes to a specified location in the chromosome (Fig. 2). After amplification, products are analyzed by agarose gel electrophoresis. Each band on the agarose gel represents a transposon insertion a given distance from the chromosomal primer site. Insertions into regions which produce significant growth defects are then represented by areas of decreased intensity on the footprinting gel. Note that either one of the two primers used for amplifying a genomic segment can also be used to analyze mutations within that segment by genomic footprinting.

As an alternative to using PCR products as substrates for *in vitro*

transposition of naturally competent organisms, a high-density insertion map of a given chromosomal region also may be obtained by performing *in vitro* transposition upon genomic DNA cloned into a vector, for example a cosmid, phage, plasmid, YAC (yeast artificial chromosome), or BAC (bacterial artificial chromosome) vector. Similar high-density mutagenesis can be performed in non-naturally competent organisms using genomic DNA cloned into an allelic replacement vector.

Lane 1 of Fig. 3A shows the analysis by agarose gel electrophoresis of the PCR products obtained from a region of the *H. influenzae* chromosome chosen for GAMBIT analysis. Areas of the gel corresponding to DNA regions that carry many *mariner* insertions contain many bands; blank regions on the gel, in contrast, correspond to segments of the chromosome that are devoid of *mariner* insertions. That the banding pattern seen in lane 1 reflects an accurate assessment of the position of insertion mutations within the targeted segment can be shown by simply moving the chromosomal primer by 114 bp (lane 2). Bands and blank regions on the gel are shifted down in migration by a distance corresponding to approximately 114 bases (molecular weights in kilobase pairs (kbp) are indicated at the right). In addition, sequencing of several gel-purified bands demonstrated that they were in the predicted loci.

GAMBIT footprinting results are quite reproducible; when two independent insertion libraries are created for a given region, the pattern exhibits only minor differences and the blank regions are unchanged (Fig. 3B, lane 3 vs. lane 4).

Fig. 3C demonstrates the use of GAMBIT to examine essential genes in the chromosome region containing a *H. influenzae* homologue of the *E. coli* gene *thyA*, which encodes thymidylate synthetase. Mutation of the *thyA* gene prevents growth on minimal medium lacking thymidine, but confers resistance

to trimethoprim. Thus, this gene provided us with the opportunity to directly test the fidelity of the system, since mutations in *thyA* can be both positively and negatively selected. A primer which hybridizes 3' to the *H. influenzae secA* gene, 5,159 bp from the *thyA* gene, was used as a chromosomal primer. When libraries selected on rich medium (BXV) are analyzed by genomic footprinting, the region corresponding to the *thyA* gene (Fig. 3C, indicated by brackets on the right) contains multiple bands. When the analysis is performed on the same mutant pool plated on a defined medium lacking thymidine (MIc), the *thyA* region PCR products are no longer seen. Since *thyA* mutants are resistant to the antibiotic trimethoprim, selection of the same pool on a medium containing trimethoprim ("Tri", 5 μg/ml) and thymidine followed by PCR analysis yields products only in the *thyA* region, confirming the identity of the bands seen in this region of the gel. Analysis of the same mutant pool with a primer which hybridizes close to the *thyA* gene demonstrates that the wide band seen in lane "Tri" can be resolved into a series of bands that correspond to multiple *mariner* inserts in the *thyA* gene (Fig. 3D).

We have found several DNA regions with a decreased number and intensity of PCR products. Some regions contained no detectable PCR products. For example, no bands could be seen in the region in *H. influenzae* corresponding to an orf with a high degree of similarity to the *E. coli* gene *surA* (Fig. 3E). In *E. coli* this gene is required for colony formation; thus, it is not surprising that insertions in *surA* are undetectable. Other regions were identified that were largely devoid of insertions but which did contain a few insertions, usually in specific reproducible locations. For example, the *H. influenzae* homologue of the *E. coli secA* gene (which encodes a portion of the preprotein translocase required for protein secretion) contained two clear insertions near the predicted 3' end of the gene (Fig. 3C, open arrowheads).

This finding is consistent with the previous observation that *E. coli* containing a truncated *secA* gene are capable of survival.

We tested whether the distribution of *mariner* insertions revealed by GAMBIT analysis reflects the essential nature of a given gene or simply site specificity of the transposon. To do this we performed *in vitro* mutagenesis of a chromosomal fragment which included the *H. influenzae secA* gene. The mutagenized DNA was then transformed into both wild-type *H. influenzae* (Rd) and an *H. influenzae* strain complemented with *E. coli secA* (RdpSecA). As discussed above, in the wild-type *H. influenzae* strain, no insertions could be found in the first 75% of the *secA* gene. However, when GAMBIT was performed on the same region in a strain complemented with *E. coli secA*, numerous transposon insertions could be found throughout the gene (Fig. 3F). These data provide strong evidence that gaps in the distribution of *mariner* insertions can be confidently attributed to the presence of an essential DNA sequence.

Using this method we studied five genomic segments in *H. influenzae* (Fig. 4) and two in *S. pneumoniae* (Table I), and identified several candidate genes required for growth or viability (Fig. 5). Many of these are known to be essential in other organisms, including *secA, surA, tmk and lgt.* Other genes have no previously known function.

Fig. 4 shows the *H. influenzae* orf analysis. As in *S. pneumoniae*, orfs with essential functions were identified using the GAMBIT/*mariner* method (Figs. 4 and 5).

An advantage of the GAMBIT technique is its ability to scan specific regions or, by more comprehensive projects, entire genomes for the presence of essential genes or DNA regions. Mutants that are reduced in growth, however, can also be detected by GAMBIT interrogation of a DNA region. Our analysis

did, in fact, detect regions with partial reductions of band intensity, suggesting that mutants with insertions in these regions had reduced the growth rates but remained viable. For example, among the genes we studied were three genes of unknown function which had been hypothesized to be members of the minimal gene set required by all bacteria. Two of these (HI0454 (see Fig. 3G) and HI1654 (not shown)) apparently cause growth attenuation when disrupted. GAMBIT analysis of HI0454 yielded detectable bands that were reduced in intensity, whereas HI1654 yielded no detectable bands. The third (HI0597), however, proved to be nonessential in *H. influenzae* under our *in vitro* conditions.

TABLE I

S.p. orf*	Position†	Essential‡	Similarity (GAP-BLAST E-value)
conserved hypothetical	840-2174	No	*Archaeoglobus fulgidus* hypo. protein, AF0170, (1e-47)
unknown	3051-3866	No	None
rbfA	4109-4459	Yes	*B. subtilis* Ribosome-binding factor A, P32731, (4e-20)
IF-2	4710-7586	Yes	*H. influenzae* Translation initiation factor IF-2, P44323, (e-153)
L7AE	7603-7902	Yes	*Enterococcus faecium* Probable ribosomal protein in L7AE family, P55768, (6e-23)
nusA	8210-9346	Yes	*B. subtilis* NusA, Z99112, (3e-96)
p15A	9390-9860	No	*B. subtilis* P15A homolog, unknown function P32726, (2e-27)
ytmQ	9995-10630	No	*B. subtilis* YtmQ, unknown function, Z99119, (5e-73)

PCR Primers used to amplify the 11,266 bp corresponding to contig 4151 of TIGR *S. pneumoniae* genomic sequence release 112197 are:
Forward 5'-CTTTCTGTAAAATGTGGGATTCAA-3' (SEQ ID NO: 1); and
Reverse 5'-AATTATTATGGAGTCGTCGTTTGG-3' (SEQ ID NO:2).
* S.p. orf designations are based on matches giving the highest GAP-BLAST score.
†Positions are given with respect to the first base of the Forward primer.
‡Essential regions as defined in the text.

GAMBIT should prove equally useful for identifying genes required for growth or viability under specific growth conditions that are more stringent than the rich *in vitro* media used exclusively here. For example, GAMBIT should allow systematic identification of the genes required by pathogenic organisms to grow and survive within a host. Fig. 6 depicts the potential outcome of such a scenario. A pool or clone of transposon-mutagenized cells is grown under conditions A and B. Condition A represents a permissive growth environment, such as rich *in vitro* growth media. Condition B represents a stringent growth environment, such as growth in a host, or growth in an *in vitro* environment that simulates a host environment, or growth in the presence of a drug at a concentration that is sub-inhibitory for wild type cells. Cells that are mutant for hypothetical gene 1 or gene 2 are viable under rich growth conditions; but only cells that are mutant for gene 2 are viable under stringent growth conditions. Therefore, gene 1 is essential for growth under stringent conditions (e.g., in a host, or in the presence of drug), but is not essential under permissive (i.e., rich growth media) conditions.

GAMBIT is well-suited to the analysis of naturally competent organisms, a group which includes important human pathogens belonging to the genera *Haemophilus, Streptococcus, Helicobacter, Neisseria, Campylobacter,* and *Bacillus*. It is also apparent that, with the use of allelic replacement vectors or efficient linear DNA transformation methods, GAMBIT should be adaptable to other bacteria and microorganisms as well. For example, the genomes of bacterial pathogens such as: *Actinobacillus actinomycetemcomitans, Borrelia burgdorferi, Chlamydia trachomatis, Enterococcus faecalis, Escherichia coli, Haemophilus influenzae, Helicobacter pylori, Legionella pneumophila, Mycobacterium avium, Mycobacterium tuberculosis, Mycoplasma genitalium, Mycoplasma pneumonia, Neisseria*

gonorrhoeae, *Neisseria meningitidis*, *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Treponema pallidum*, and *Vibrio cholerae* are either partially or entirely sequenced. Such sequence information makes possible the use of GAMBIT for the identification of drug target genes in these organisms. Drug target genes may be exploited in screening assays for the identification and isolation of antimicrobial compounds.

In addition, promoters from essential genes identified by GAMBIT, when fused to reporter genes, may be used in sensitive high-throughput screens for the identification of compounds that decrease expression of essential genes at the transcriptional or post-transcriptional stages. Such screens are useful for the detection of antimicrobial compounds. Analogous screens for compounds that *increase* expression of essential genes also are useful, for example, for identifying compounds that increase expression of a gene that promotes survival (e.g., an anti-apoptotic gene) in an animal or plant cell. Alternatively, increased or decreased expression of essential genes identified by GAMBIT can be detected by other methods known to skilled artisans, such as by PCR or ELISA. In either case, the assays utilize standard molecular and cell biological techniques known to those skilled in the art. Such assays are readily adaptable to high-throughout screening assays for identifying or isolating novel compounds that regulate expression of essential genes.

Test Compounds and Extracts

In general, compounds are identified from large libraries of both natural product and synthetic (or semi-synthetic) extracts or chemical libraries according to methods known in the art. Those skilled in the field of drug discovery and development will understand that the precise source of test extracts or compounds is not critical to the screening procedure(s) of the

invention. Accordingly, virtually any number of chemical extracts or compounds can be screened using the methods described herein. Examples of such extracts or compounds include, but are not limited to, plant-, fungal-, prokaryotic- or animal-based extracts, fermentation broths, and synthetic compounds, as well as modification of existing compounds. Numerous methods are also available for generating random or directed synthesis (e.g., semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. Synthetic compound libraries are commercially available from Brandon Associates (Merrimack, NH) and Aldrich Chemical (Milwaukee, WI). Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant, and animal extracts are commercially available from a number of sources, including Biotics (Sussex, UK), Xenova (Slough, UK), Harbor Branch Oceangraphics Institute (Ft. Pierce, FL), and PharmaMar, U.S.A. (Cambridge, MA). In addition, natural and synthetically produced libraries are produced, if desired, according to methods known in the art, e.g., by standard extraction and fractionation methods. Furthermore, if desired, any library or compound is readily modified using standard chemical, physical, or biochemical methods.

In addition, those skilled in the art of drug discovery and development readily understand that methods for dereplication (e.g., taxonomic dereplication, biological dereplication, and chemical dereplication, or any combination thereof) or the elimination of replicates or repeats of materials already known for their anti-pathogenic activity should be employed whenever possible.

When a crude extract is found to have a desired modulating activity, or a binding activity, further fractionation of the positive lead extract is necessary to isolate chemical constituents responsible for the observed effect.

Thus, the goal of the extraction, fractionation, and purification process is the careful characterization and identification of a chemical entity within the crude extract having the desired activity. Methods of fractionation and purification of such heterogenous extracts are known in the art. If desired, compounds shown to be useful agents for the treatment of pathogenicity are chemically modified according to methods known in the art.

Uses

For therapeutic uses, the compounds, compositions, or agents identified using the methods disclosed herein may be administered systemically, for example, formulated in a pharmaceutically-acceptable buffer such as physiological saline. Treatment may be accomplished directly, e.g., by treating the animal with antagonists which disrupt, suppress, attenuate, or neutralize the biological events associated with a pathogen. Preferable routes of administration include, for example, inhalation or subcutaneous, intravenous, interperitoneally, intramuscular, or intradermal injections which provide continuous, sustained levels of the drug in the patient. Treatment of human patients or other animals will be carried out using a therapeutically effective amount of an anti-bacterial agent in a physiologically-acceptable carrier. Suitable carriers and their formulation are described, for example, in Remington's *Pharmaceutical Sciences* by E.W. Martin. The amount of the anti-bacterial agent to be administered varies depending upon the manner of administration, the age and body weight of the patient, and with the type of disease and extensiveness of the disease. Generally, amounts will be in the range of those used for other agents used in the treatment of other microbial diseases, although in certain instances lower amounts will be needed because of the increased specificity of the compound. A compound is administered at a

dosage that inhibits microbial proliferation or survival. For example, for systemic administration a compound is administered typically in the range of 0.1 ng - 10 g/kg body weight.

For agricultural uses, the compounds, compositions, or agents identified using the methods disclosed herein may be used as chemicals applied as sprays or dusts on the foliage of plants, or in irrigation systems. Typically, such agents are to be administered on the surface of the plant in advance of the pathogen in order to prevent infection. Seeds, bulbs, roots, tubers, and corms are also treated to prevent pathogenic attack after planting by controlling pathogens carried on them or existing in the soil at the planting site. Soil to be planted with vegetables, ornamentals, shrubs, or trees can also be treated with chemical fumigants for control of a variety of microbial pathogens. Treatment is preferably done several days or weeks before planting. The chemicals can be applied by either a mechanized route, e.g., a tractor or with hand applications. In addition, chemicals identified using the methods of the assay can be used as disinfectants.

In addition, the antipathogenic agent may be added to materials used to make catheters, including but not limited to intravenous, urinary, intraperitoneal, ventricular, spinal and surgical drainage catheters, in order to prevent colonization and systemic seeding by potential pathogens. Similarly, the antipathogenic agent may be added to the materials that constitute various surgical prostheses and to dentures to prevent colonization by pathogens and thereby prevent more serious invasive infection or systemic seeding by pathogens.

Methods

Bacterial Culture

H. influenzae Rd strain (ATCC #9008) (J. Reidl and J. J. Mekalanos; *J. Exp. Med.* 183: 621-629 (1996)), the gift of Andrew Wright, was grown on BHI medium supplemented with 5% Levinthal's base (BXV) (H. Alexander, in: Bacterial and Mycotic Infections of Man, R. Dubos, J. Hirsch, Eds. (JB Lipincott, Philadelphia, 1965), vol. 724-741) or on MIc medium (R. M. Herriott, E. M. Meyer, M. Vogt, *J. Bacteriol.* 101: 517-524 (1970)).

S. pneumoniae (strain Rx1) (N. B. Shoemaker and W. R. Guild, *Mol. Gen. Genet.* 128: 283-290 (1974)) was grown on tryptic soy agar supplemented with 5% defibrinated sheep blood.

In Vitro Transposition

Minitransposons were constructed which contained the inverted repeats of the *Himar* transposon and ~100 bp of *Himar* transposon sequence flanking either a kanamycin resistance gene (M. F. Alexeyev, I. N. Shokolenko, T. P. Croughan. *Gene* 160: 63-67 (1995)) for *H. influenzae* or a chloramphenicol resistance gene (J. P. Claverys, A. Dintilhac, E. V. Pestova, B. Martin, D. A. Morrison. *Gene* 164: 123-128 (1995)) for *S. pneumoniae*. Transposition reactions were performed using purified *Himar* transposase as previously described (D. J. Lampe, *supra*; herein incorporated by reference).

Templates for transposition were either chromosomal DNA or PCR products. PCR of ~10 kb chromosomal regions was performed using *Taq* polymerase (Takara) and *Pfu* polymerase (Stratagene) at a 10:1 ratio, 100 pmol of primers and 30 cycles of amplification (30 seconds denaturation at 95°C, 30 seconds annealing at 62°C and 5 minutes extension at 68°C with 15 seconds added to the extension time for each cycle). Gaps in transposition products

were repaired with T4 DNA polymerase and nucleotides followed by T4 DNA ligase with ATP (New England Biolabs) (J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular Cloning-A Laboratory Manual, Second Edition, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989)).

Repaired transposition products were transformed into *H. influenzae* as previously described (G. J. Barcak, M. S. Chandler, R. J. Redfield, J. F. Tomb, *Meth. Enzymol.* 204:321-342 (1991)). and into *S. pneumoniae* as previously described using CSP-1 for competence induction (L. S. Havarstein, G. Coomaraswamy, D. A. Morrison; *Proc. Natl. Acad. Sci.* USA. 92: 11140-11144 (1995)).

Genomic Footprinting

Genomic footprinting was carried out as described (I. R. Singh, R. A. Crowley, P. O. Brown, *Proc. Natl. Acad. Sci.* USA. 94: 1304-9, 1997; herein incorporated by reference) using a transposon-specific primer (5'-CCGGGGACTTATCAGCCAACC-3'; SEQ ID NO: 3) and primers specific to each chromosomal region designed using chromosomal sequence from The Institute for Genomic Research (TIGR). The chromosomal primers for the experiments shown in Figs. 3A-3G lie within or near the following loci (TIGR designation):

a) HI0449 (primer in lane 1 (5'-CGCCTTTTTGTAAATCACGCATCGC-3'; SEQ ID NO: 4) hybridizes 114 bp 5' of the primer in lane 2 (5'-GCGGATGAAACAAA TCGACCAGCAG-3'; SEQ ID NO: 5));

b) HI1658 (5'-TCACGCCGCTGATTTTGCTGG-3'; SEQ ID NO: 6);

c) HI0911 (5'-GGGAGCAAGAAAAGCGACAGAAGCC-3'; SEQ ID NO: 7);

d) HI0905 (5'-AAATCATCCATCGTGACCCA-3'; SEQ ID NO: 8);

e) HI0461 (5'-CCCGAATAAATTGCTTATCGCCTCG-3'; SEQ ID NO: 9);

f) HI0911 (5'-GGGAGCAAGAAAAGCGACAGAAGCC-3'; SEQ ID NO: 10); and

g) HI0456 (5'-CAGGCGTATCAGGGTGGTGGACG-3'; SEQ ID NO: 11).

PCR was performed using the protocol described above. Potential *S. pneumoniae* orfs were analyzed for homology using the GAP-BLAST program (S. F. Altshul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Nucleic Acids Res.* 25: 3389-3402, 1997).

PCR products were analyzed by gel electrophoresis on 0.8% agarose gels. Plasmid pSecA, which contains the *E. coli secA* gene, was constructed by cloning the *Bam*HI fragment from pT7secA (M. G. Schmidt and D. B. Oliver; *J. Bacteriol.* 171: 643-9 (1989)), the gift of Carol Kumamoto, into the *Bgl*II site of the *E. coli-H. influenzae* shuttle plasmid pGJB103 (G. J. Barcak, M. S. Chandler, R. J. Redfield, J. F. Tomb, *Meth. Enzymol.* 204:321-42 (1991)), the gift of Gerard Barcak.

Isolation of Conditional Mutations in Essential Genes

Isolation of conditional mutations in essential genes represents a powerful next step in characterization of genes identified by GAMBIT. Temperature sensitive mutations are a class of functional mutations in protein coding regions that allow depletion of the active form of the non-permissive temperature.

We have begun analysis of essential genes identified by GAMBIT by isolating temperature sensitive mutations. Briefly, DNA containing a mariner insertion near an essential gene is amplified by mutagenic PCR (using standard PCR conditions modified by the addition of 125μM $MnCl_2$ to the reaction) and transformed into *H. influenzae*. This mutagenesis method allows nucleotide misincorporation during amplification and is predicted to give a relatively high

proportion of missence mutations in comparison with methods which induce DNA damage, such as UV irradiation, which leads to relatively high frequency of deletion mutations. In addition, since DNA damage is not generated by this procedure, second site mutations due to the induction of DNA repair mechanisms of the host cell are absent or greatly reduced in frequency.

H. influenzae transformants are selected on kanamycin and screened for growth at 30°C and lack of growth at 37°C. The mutation is then mapped by rescuing growth at the non-permissive temperature via transformation with PCR products corresponding to the wild-type region being analyzed. By transforming with wild-type DNA it is possible to map the mutation to a specific open-reading frame. If necessary, further mapping can be accomplished by sequencing the mutant allele. Using this method we have isolated conditional lethal mutations in the *H. influenzae secA* homologue and in a conserved gene.

This set of techniques provides a rapid way to confirm essentiality and characterize genes identified by GAMBIT. The linked insertions generated by GAMBIT near each essential gene automatically provide the starting material for these experiments. Since cloning in recombinant plasmids is not necessary in naturally competent organisms, the method eliminates time-consuming steps that would be needed to generate complementing clones. At the same time, the method provides a strain in which the gene of interest can be selectively, and inducible depleted from the cell.

Conditional mutations of this kind can be used to further define the functions of essential genes. In addition, conditional mutations in essential genes can be used to produce cells with intermediate levels of the essential protein. These mutant may be used for drug sensitivity screens.

Oher Embodiments

All publications mentioned in this specification are herein incorporatcd by reference to the samc extent as if each independent publication was specifically and individually indicated to be incorporated by reference.

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications. This application is intended to cover any variations, uses, or adaptations following, in general, the principles of the invention and including such departures from the prescnt disclosure within known or customary practice within the art to which the invention pertains and may be applied to the essential features hereinbefore set forth, and follows in the scope of the appended claims.

What is claimed is:

1. A method for locating an essential region in a portion of DNA from the genome of an organism, said method comprising:

a) mutagenizing DNA having the sequence of said portion of DNA, said mutagenizing using *in vitro* mutagenesis with a transposon;

b) transforming cells of said organism with the mutagenized DNA of step a);

c) identifying cells containing said mutagenized DNA; and

d) locating said essential region of said portion by detecting the absence of transposons in said region in said mutagenized cells containing said mutagenized DNA.

2. The method of claim 1, wherein said portion of DNA is amplified by PCR prior to said mutagenesis.

3. The method of claim 1, wherein said portion of DNA is cloned into a vector prior to said *in vitro* transposon mutagenesis.

4. The method of claim 1, wherein said transposon contains a selectable marker.

5. The method of claim 1, wherein said transposon is *mariner*.

6. The method of claim 5, where said method further comprises the use of *Himar 1* transposase.

7. The method of claim 1, wherein said locating of an essential region is done by performing PCR footprinting on a pool of transposon-

mutagenized cells, wherein said PCR is performed using a primer that hybridizes to said transposon, plus a primer that hybridizes to a specific location on said chromosome, and wherein the products of said PCR are separated on a footprinting gel, wherein a PCR product on said gel represents a region of said chromosome that does not contain an essential gene, and wherein the lack of said PCR product in an area of said gel, where said PCR product is expected, represents a region of said chromosome that contains an essential gene, or, wherein a low level of said PCR product on said gel, relative to other PCR products on said gel, represents a region of said chromosome that contains an essential gene.

8. The method of claim 1, wherein prior to said transforming, said mutagenized DNA is subjected to gap repair using DNA polymerase and DNA ligase.

9. The method of claim 1, wherein said cell has a haploid growth phase.

10. The method of claim 1, wherein said cell is a single-cell microorganism.

11. The method of claim 1, wherein said cell is naturally competent for transformation.

12. The method of claim 1, wherein said cell is made competent for transformation.

13. The method of claim 1, wherein said cell is a fungus.

14. The method of claim 13, wherein said fungus is a yeast.

15. The method of claim 14, wherein said yeast is *Saccharomyces cerevisiae*.

16. The method of claim 10, wherein said microorganism is a bacterium.

17. The method of claim 16, wherein said bacterium is a gram-positive bacterium.

18. The method of claim 17, wherein said bacterium is selected from the group consisting of: *Actinobacillus actinomycetemcomitans*; *Borrelia burgdorferi; Chlamydia trachomatis; Enterococcus faecalis; Escherichia coli; Haemophilus influenzae; Helicobacter pylori; Legionella pneumophila*; *Mycobacterium avium; Mycobacterium tuberculosis*; *Mycoplasma genitalium*; *Mycoplasma pneumonia*; *Neisseria gonorrhoeae*; *Neisseria meningitidis*; *Staphylococcus aureus*; *Streptococcus pneumoniae*; *Streptococcus pyogenes*; *Treponema pallidum*; and *Vibrio cholerae*.

19. The method of claim 1, wherein said transposon-mutagenized DNA is recombined into said chromosome using an allelic replacement vector.

20. The method of claim 1, wherein said transposon contains a selectable marker gene, and wherein said identifying said cells containing said

mutagenizcd DNA is based upon the ability of said cells to grow on selective medium, wherein a cell containing a transposon can grow on said selective medium, and a cell lacking a transposon cannot grow, or grows more slowly, on said selective medium.

21. The method of claim 1, wherein said transposon contains a reporter gene, wherein said identifying of said cells containing said mutagenized DNA is based on a reporter gene assay, wherein a cell confirming a transposon expresses said reporter gene and a cell lacking a transposon does not express said reporter gene.

22. The method of claim 1, wherein said *in vitro* mutagenesis is high saturation mutagenesis.

23. A method for isolating a compound that modulates the expression of a nucleic acid sequence operably linked to a gene promoter, said method comprising:

a) providing a cell expressing a nucleic acid sequence operably linked to a gene promoter, wherein said gene promoter is the gene promoter for: HI0455; HI0456; HI0458; HI0599; HI0887; HI0904; HI0906; HI0907; HI0908; HI0909; HI1650; HI1651; HI1654; HI1655; *S. pneumoniae* rbfA; *S. pneumoniae* IF-2; *S. pneumoniae* L7AE; or *S. pneumoniae* nusA;

b) contacting said cell with a candidate compound; and

c) detecting or measuring expression of said gene following contact of the cell with said candidate compound.

24. A method for identifying a nucleic acid sequencc that is

essential for cell growth or viability, said method comprising:

a) expressing in a cell (i) a first nucleic acid sequence operably linked to a gene promoter, wherein said gene promoter is the gene promoter for: HI0455; HI0456; HI0458; HI0599; HI0887; HI0904; HI0906; HI0907; HI0908; HI0909; HI1650; HI1651; HI1654; HI1655; *S. pneumoniae* rbfA; *S. pneumoniae* IF-2; *S. pneumoniae* L7AE; or *S. pneumoniae* nusA; and (ii) a second nucleic acid sequence; and

b) monitoring the expression of said first nucleic acid sequence, wherein an increase in said expression identifies said second nucleic acid sequence as being essential for cell growth or viability.

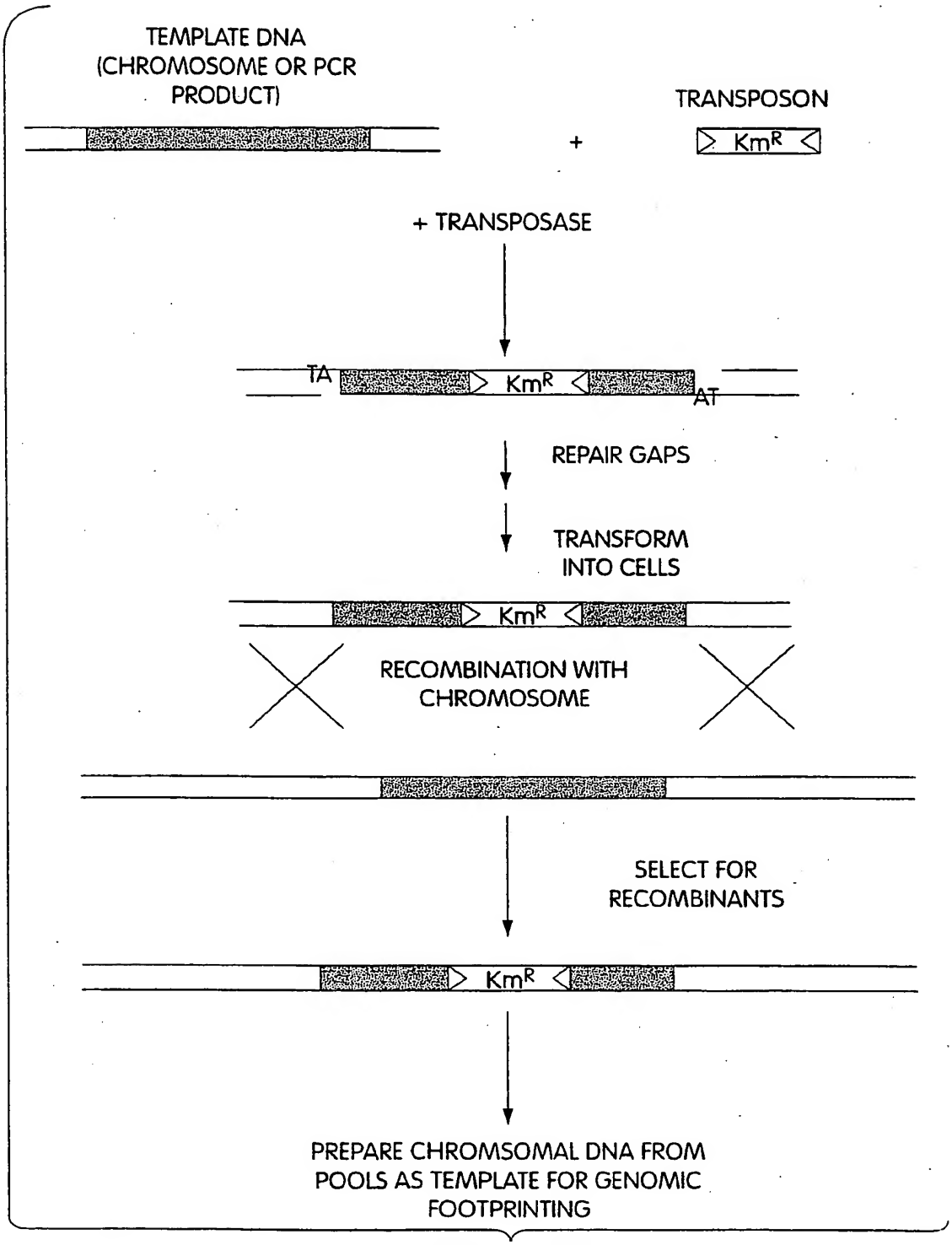
TEMPLATE DNA
(CHROMOSOME OR PCR
PRODUCT)
TRANSPOSON
+
KmR
+ TRANSPOSASE
TA
KmR
AT
REPAIR GAPS
TRANSFORM
INTO CELLS
KmR
RECOMBINATION WITH
CHROMOSOME
SELECT FOR
RECOMBINANTS
KmR
PREPARE CHROMSOMAL DNA FROM
POOLS AS TEMPLATE FOR GENOMIC
FOOTPRINTING

Fig. 1A

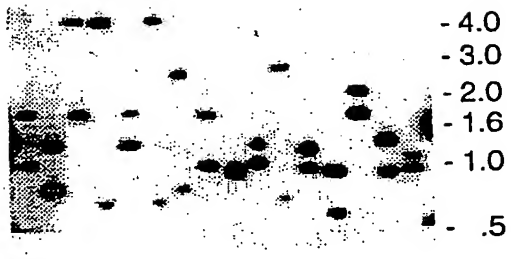

Fig. 1B

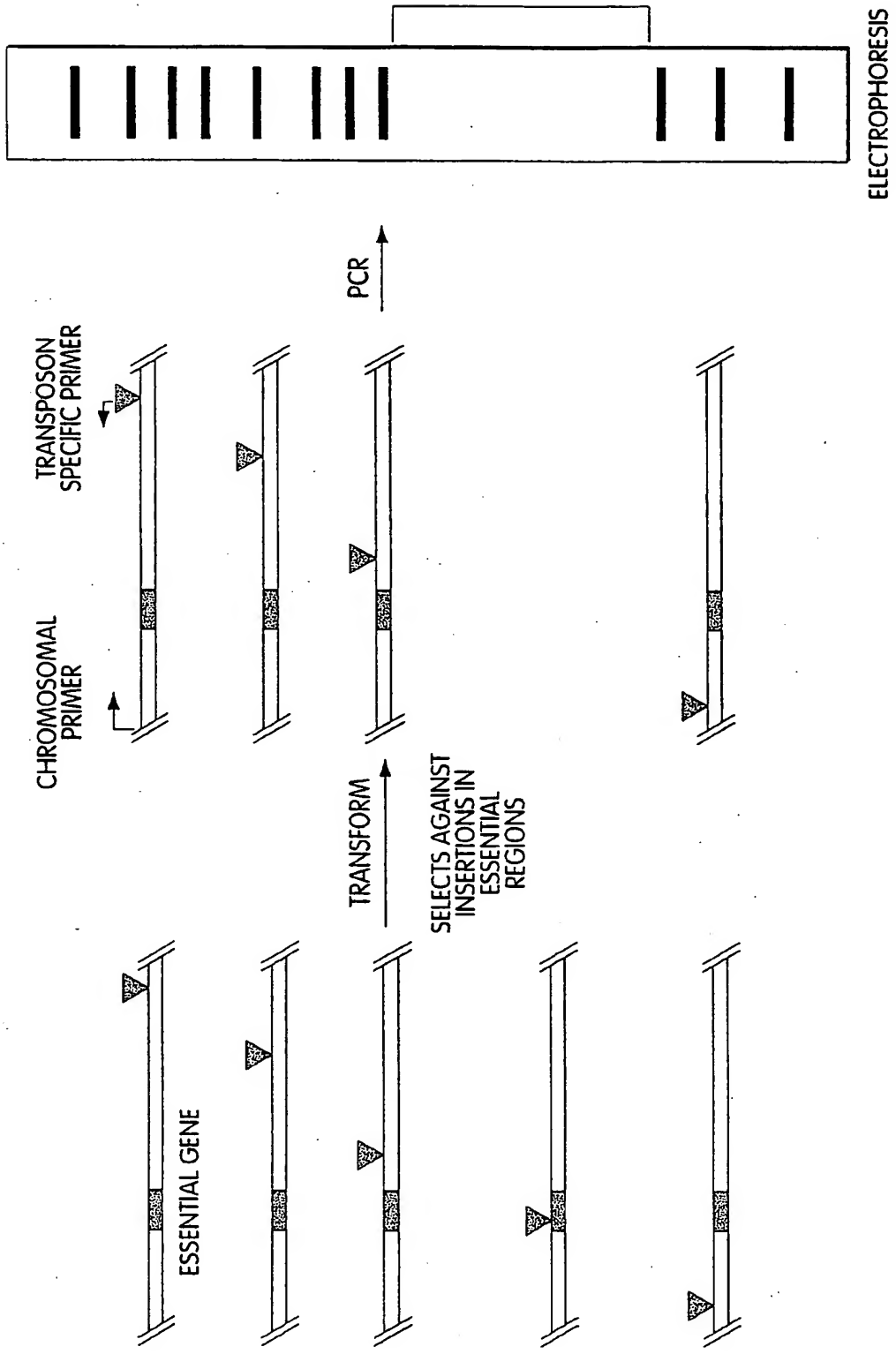
ELECTROPHORESIS
PCR
TRANSPOSON SPECIFIC PRIMER
CHROMOSOMAL PRIMER
TRANSFORM
SELECTS AGAINST INSERTIONS IN ESSENTIAL REGIONS
ESSENTIAL GENE

Fig. 2

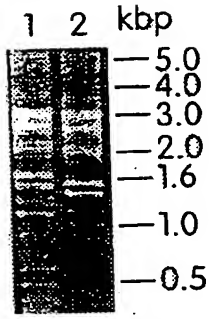

Fig. 3A

Fig. 3B

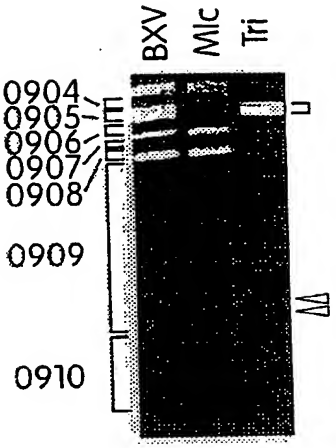

Fig. 3C

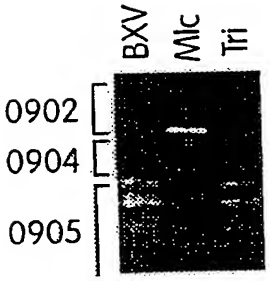

Fig. 3D

Fig. 3E

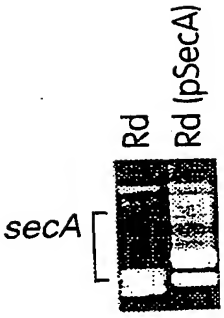

Fig. 3F

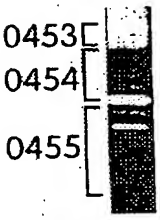

Fig. 3G

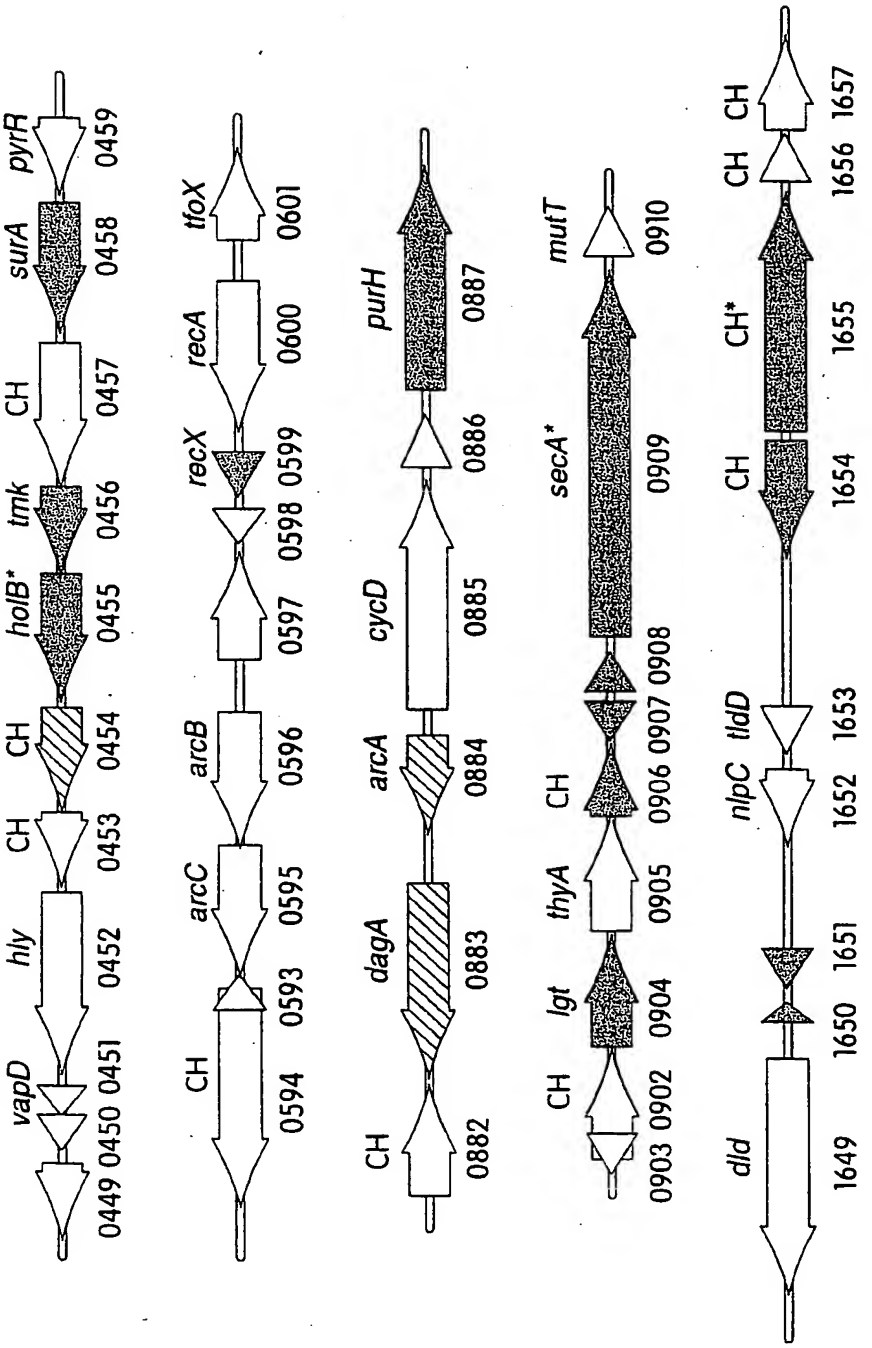
vapD hly CH CH holB* tmk CH surA pyrR
0449 0450 0451 0452 0453 0454 0455 0456 0457 0458 0459
CH arcC arcB recX recA tfoX
0594 0593 0595 0596 0597 0598 0599 0600 0601
CH dagA arcA cycD purH
0882 0883 0884 0885 0886 0887
CH lgt thyA CH secA* mutT
0903 0902 0904 0905 0906 0907 0908 0909 0910
dld nlpC tldD CH CH* CH CH
1649 1650 1651 1652 1653 1654 1655 1656 1657

Fig. 4

HI0455

ATGACCGCACTTTACCCTTGGCTAATGCCAATTTATCATCAAATTGCTCAAACCTTTGACGAAGGGTTGG
GGCATCATGCTGTGCTGATTAAAGCTGATTCTGGTTTAGGCGTAGAGAGTTTATTTAATGCACTTGCACA
GAAAATAATGTGTGTAGCTCAAGGCGATAAACCTTGTGGTCAATGCCATTCTTGTCATTTAATGCAAGCC
CATAGCCATCCAGATTATCACGAATTAAGCCCCATTAACGGTAAGGATATTGGCGTTGATCAAGTACGCG
ACATTAATGAAATGGTTGCGCAGCACGCACAACAAAACGGCAATAAAGTGGTGTATGTGCAAGGGGCGGA
ACGTTTAACGGAAGCGGCTGCTAATGCATTATTGAAAACATTGGAAGAGCCTCGTCCAAATACTTATTTT
TTACTTCAAGCGGATAGTTCGGCAAGTTTGTTAGCAACTATTTACAGTCGATGCCAAGTGTGGAATCTTT
CCGTGCCTAATGAAGAAATTGCTTTTGAATGGTTAAAATCAAAAAGTGCGGTAGAAAATCAGGAAATTTT
GACCGCACTTGCGATGAATCTTGGGCGTCCGCTTTTAGCATTAGAAACGTTACAAGAAGGATTTATTGAA
CAGCGTAAAAACTTCTTACGTCAATTTTGGGTGTTCTATCGCCGACGTTCGCCATTGGAATTGCTTCCGT
TGTTTGATAAAGAACGCTATGTTCAGCAAGTGGATTGGATTTTGGCTTTTCTTTCTGATTGTTTAAAACA
TAAACTTGAAATTGATAGTCATCGACAAGTGGCTGATCTTGGCCGTGGTATCGAACAATTCAGCGACGAG
CAAACTGCCCTTGGTTTATTACAAGCCATTAAAATTATGCAAAAAGTGCGGTCAGATTTGCTTACAATTA
ATGGTGTGAATGTTGAATTAATGCTATTGGATGGCTTGACACGATTAGTCACAGAAGTATTTGAAACGCA
ATAA

SEQ ID NO:12

Fig. 5A

>HI0456
ATGAAAGGAAAGTTTATTGTCATTGAGGGCTTAGAAGGTGCGGGGAAAAGCTCCGCTCATCAGTCTGTAG
TGCGAGTTTTTGCATGAACTTGGTATTCAAGATGTTGTGTTTACGCGCGAGCCTGGTGGAACGCCACTGGC
TGAGAAATTACGTCATCTCATCAAACATGAAACCGAAGAACCCGTGACAGATAAAGCAGAGTTATTAATG
CTTTATGCGGCTCGTATTCAGTTGGTGGAAAATGTGATTAAACCTGCTTTAATGCAAGGGAAATGGGTAG
TGGGCGATCGTCACGATATGTCATCTCAGGCGTATCAGGGTGGTGGACGTCAATTAGACCCGCATTTTAT
GCTCACCTTGAAAGAAACCGTATTAGGTAATTTTTGAGCCAGATCTCACACAATTTATTTGGATATAGATCCG
AGCGTCGGTTTAGCGCGAGCTCGTGGACGTGGCGAGTTAGATCGTATTGAGCAAATGGATTTAGATTTTT
TCCATCGTACTCGAGCACGCTATTTAGAATTAGTAAAAGATAATCCCAAAGCAGTGGTGATTAATGCAGA
GCAGAGTATTGAACTTGTTCAAGCTGATATTGAAAGTGCGGTAAAAAATTGGTGGAAATCAAACGAAAAA
TGA

SEQ ID NO:13

Fig. 5B

>HI0458
ATGAAAATGAAAAAATTTGTTTTAAGATCTTTTTTATTGGCTACTTTAGGTTGTGTTGCTTTTACTTCTA
TGGCACAAGCGGAGGAACGTGTCGTAGCAACAGTGGATGGTATTCCTGTTTTAGAAAGTCAAGTGCGTGC
CAATATGGGTAAAAAAGGTGATCGCCAAAGTGCGATTGATAAAATTATTGATGATATTTTGGTGCAAAAA
GCAGTTCAAGAATCGGGAGTCAAAATTGATCCGCGTGAAATTGATCATATTGTGGAAGATACCGCAGCTA
GAAATGGTTTAACTTATGGTCAATTTTTGGATGCGTTAGATTATCAAGGCATTTCATTAAATACATTCCG
TCAGCAAATTGCCAATCAAATGGTGATGGGGGCTGTACGTAACAAAGCTATTCAAGAAAGCATTGATGTA
ACGCGTGAAGAAGTTGTCGCACTTGGTCAAAAAATGTTGGATGAGGCAAAATCACAAGGCACTGCACAAA
AAGTTACAGGTAAAGAATACGAAGTGCGTCACATTTTGTTAAAACTTAATCCATTGTTAAATGATGCTCA
AGCAAAAAAACAATTAGCTAAAATTCGTTCTGATATTATTGCAGGTAAAACAACTTTCGCTGATGCCGCA
TTAAAATATTCTAAAGATTATTTATCGGGTGCGAATGGCGGTAGTTTAGGTTATGCGTTCCCAGAAACTT
ATGCACCACAGTTTGCACAAACCGTCGTGAAAAGTAAACAAGGTGTGATTTCTGCACCATTTAAAACTGA
GTTTGGTTGGCATATTTTGGAAGTAACTGGCGTACGTGATGGCGATCTTACAGCAGAAGCCTACACACAA
AAAGCATATGAACGTTTAGTAAATACTCAATTACAAGATGCGACGAACGATTGGGTTAAAGCATTGCGTA
AAAGAGCGAATATTCAGTATTTTAATAAATAACAAATTCATCGCTACGAAATCCGTGTGATAGCATATAT
TGCCAAATTTTTTGTTTCATTTTAG

SEQ ID NO:14

Fig. 5C

>HI0599
ATGCAGGAAAAGAACTTTTCGGAAGAAATTGATGCATTATCTCGATGCCAAGCAAAAATTGGC
AAAGTGATCGTCGTTTTTCAGAAAATTATCTAAATTCACGCGTGCAAAAAGGTTATGGTGTAGGAAGAAT
TCGACAAGAATTACGCCAATTAAAAGGTGTGTCTTCTGATATTATTGATGAAGTTTTAATGGAATCAGAA
ATTGATTGGTATGAAATGGCTGAGAACTTGTTACGTAAAAAATTCCCAAATTATAACGAACAGCAAACGC
CTAAAATGAAACAAAAAATTTGGCAATATATGCTATCACACGGATTTCGTAGCGATGAATTTGCTGATTT
AATTGGGCAAAACCAAAGTGAATGGGATTAA

SEQ ID NO:15

Fig. 5D

>HI0887

ATGGCAGATCGTCCAATTCGTCAGGCTTTACTGAGTGTGTCTGATAAAACGGGTATTGTAGAGTTTGCTC
AAGGTTTAGTTAAACGTGGTGTAAAACTACTTTCAACAGGTGGAACGGCAAAACTTTTAGCACAAAATGC
TTTACCTGTAATAGAAGTGTCTGATTACACAGGTTTCCCAGAAATGATGGACGGTCGAGTGAAAACCTTA
CATCCCAAAGTACATGGCGGTATTCTTGGTCGTCGTGGTACAGATGATGCCATCATGCAGCAACATGGCA
TTGAAGGCATTGATATGGTCGTTGTGAATTTATATCCCTTTGCTGCCACTGTGGCAAAACCTGATTGCAC
TTTGGCTGATGCGGTAGAAAATATCGATATTGGGGGGCCTACAATGGTGCGTTCTGCAGCGAAAAACCAC
AAAGATGTAGCGATCGTGGTTAATAATCATGATTTCAACGCAATTCTAGCCGAAATGGATCAACATCAAA
ACAGCCTAACTTTTGAAACTCGTTTTGACCTTGCGATTAAAGCATTTGAACATACCGCTCAATATGATTC
TATGATTGCCAACTATTTCGGTCAGCTAGTAAAACCTTATCATATTGCAGAGGAAGAAGAGGCGAATGCG
AAGTGCGGTCAATTCCCACGGACTTTAAACCTTAACTTCGTGCGTAAACAAGCTATGCGTTACGGCGAAA
ACTCCCATCAAAATGCGGCATTTTATGTTGATTTAAATGTGAAAGAAGCGAGCGTGGCTACAGCTAATCA
ACTGCAAGGTAAAGCCTTGTCTTACAATAATATTGCCGACACTGATGCAGCACTTGAATGCGTGAAAGAA
TTTGACGATCCAGCTTGCGTAATCGTTAAACATGCCAATCCATGTGGTGTGGCGTTAGGTAAGGATATTT
TAGACGCTTATAATCGCGCTTACCAAACCGATCCAACGTCTGCTTTTGGCGGCATTATTGCTTTTAACCG
TGAATTAGACGAAAAAACGGCGAATGAAATTGTGGAACGCCAATTCGTTGAAGTGATTATCGCACCGAAA
GTTTCTGCTGAAGCGCAAGAAGTAATGAAGCGTAAGAAAAATGTGCGTTTGCTTGAATGTGGTGAATGGA
CTTCTCGTTCCGAACGTTTGGATTTCAAACGTGTAAACGGCGGTTTATTAGTACAAGATGCGGATTTAGG
TATGGTTGGCGTGGATGATTTAAAAGTCGTGAGTAAACGTCAGCCAACTGAACAAGAATTAAAAGACTTA
TTATTCTGCTGGAAAGTGGCAAAATTTGTGAAATCGAATGCCATTGTTTACGCCAAAGACAATCAAACTA
TCGGCATTGGTGCAGGTCAAATGAGCCGCGTATATTCTGCGAAGATTGCGGGTATTAAAGCGCAAGATGA
AGGTTTAGAAGTGGCTGGTTGTGTGATGGCATCTGATGCGTTCTTCCCATTCCGTGACGGCATTGATGCG
GCGGCGAAAGTGGGTATTCAATGTGTGATCCATCCAGGTGGATCAATGCGCGATCAAGAAGTCATTGATG
CGGCGGATGAACATAATATGGTAATGGTATTGACTGGAATGCGACATTTTAGACATTAA

SEQ ID NO:16

Fig. 5E

>HI0904

ATGAATTCAAATTATTTACTTCTTCCCCACTTTGATCCGAGTATTTTTACGCTTGGCGATAGTAATATCG

GCTTACGTTGGTATGGCTTGATGTACCTTTTAGGTTTTTGTTTTTGCACGTTGGCTTGCGGTTCGCCGTGC

TAATCGCCCAAATAGCGGTTGGACAGTAGATCAAGTTGATAGCTTACTTTTCAACGGTTTTATGGGGGTG

TTTATTGGCGGACGTGTTGGCGATGTATTTTTCTATAATCTCGATCATTTCTTACAAGAACCACTTTATT

TATTCCGCGTTTGGGAAGGTGGAATGTCGTTCCACGGTGGCTTAATTGGTGTAATTGTTGCTATGATTTG

GACATCTTATTCTCAAAAACGTAATTTTTTGGCAAACGGCTGATTTTGTTGCGCCTTTGATTCCGTTTGGT

TTAGGTTTAGGCAGAATTGGTAATTTCATTAATCTTGAACTATGGGGACGCGAAACGAATGTGCCTTGGG

CAATGATTTTTCCCGAATGATCCTCTTTTACTGCCTCGTCATCCATCACACTTTTATGAAGCCTTTTTTAGA

AGGCCTGGTGTTGTTTACGATTCTGAATATTTTTTATTAAAAAACCACGTCCAATGGCTTCTGTTGCAGGT

TTATTCTTAATTGGTTATGGCGTCTTCCGTTTTATTGTGGAATATGTGCGTGAACCTGAAGTTGAAAATT

TCTTTGGGATTATTACACGAGGGCAAGCCCTTTGCTTGCCGATGATTATTGGTGGTGCTTTCATTATGGC

TTGGGCTTATTCACGCAAAAGTGCGGTCATAAAATAG

SEQ ID NO:17

Fig. 5F

>HI0906
ATGGATGCAGCAAAAGTGCGGTCAGAATTTGACGAGAAAATGATGCGCTACGCCCTTGAGCTTGCCGATA
AAGCGGAAGCGTTAGGTGAGATTCCTGTGGGGGCGGTGCTGGTGGATGACGCTAGAAATATTATTGGAGA
AGGTTGGAATCTCTCCATTGTTCAAAGTGATCCTACTGCACATGCTGAAATTATCGCTTTGCGTAATGGT
GCGAAAAATATTCAAAATTATCGCCTACTGAATAGCACGCTTTATGTGACATTAGAGCCTTGCACAATGT
GCGCAGGGGCAATTTTGCATAGCCGTATTAAACGTCTTGTGTTCGGTGCATCTGATTATAAAACTGGCGC
GATTGGATCACGTTTTCATTTTTTTGATGATTACAAGATGAATCATACTTTAGAGGTTACATCTGGCGTA
TTGGCAGAAGAGTGTAGTCAAAAATTGAGTACATTTTTTCAGAAAAGACGCGAGGAGAAAAAAATAGAGA
AAGCATTATTAAAAAGTCTGAGTGATAAGTAA

SEQ ID NO:18

Fig. 5G

>HI0907
ATGGAAAACAAGGCTGAGCGTTATCAAAAAGCAGTCAATATTACGGATGTGCTTGAGCAATCGCCCTTTG
CCAAAATAATCAAAAAAGGTCTTGCTATCAATGAAATCAATCAAAAATTTAACCGCATTTTTCCACAGGA
ATTTCACGGCAAATTTCGTATTGGTAATATGACAGATAACTCAATTTTTATTGAGACAGCAAATGCGATC
GTTCGCCAAGGAATTTTATTCAGACAGACAGAATTGTTGAAACTCATTCAAGAAGAGTTTCCGCAAGTAA
CAGGATTTGAGATAACGATCAATCCTGGATTTTAA

SEQ ID NO:19

Fig. 5H

>HI0908
ATGATTTCAGGCACTGTCAAACCGAATTTTTGGTCGCGATTACTTTTAAGTATCATCGCAATTTTTGCTT
TGCCTAACGCACACAAAGTTTTGAAAATCAAAATAATACGGAAAATTATTCCTCAAGTGTTTCCATTCAACA
AGCGTTAGAAACGGTAAAAGTTGCTCGTGAAGTGCAACGACAAGCCATTCCTCAACCTTCAATTTCCCGT
CAAACTGAAAAACAACTTAAAATTCAACCGCACTTTTTTACTGAAGCGTTGAATATTAGCGCGCCAATTC
GAGCAGGCCCCTTGCTTATTTAA

SEQ ID NO: 20

Fig. 5I

>HI0909

ATGAGCATTTTAACAAGAATTTTTGGTAGTCGTAATGAACGCGTTTTACGTAAATTAAAAAAACAAGTCG
TAAAAATTAATAAAATGGAGCCTGCTTTTGAGGCATTAAGTGATGATGAATTAAAAGCAAAAACACAAGA
GTTTCGTGATCGTTTAAGTGGTGGCGAAACTTTGCAACAAATTTTACCAGAAGCATTCGCAACGGTACGC
GAAGCAAGTAAGCGTGTGCTTGGTATGCGCCATTTTGATGTTCAGCTTATCGGTGGGATGGTATTGACTA
ACCGCTGTATCGCAGAAATGCGTACTGGTGAAGGTAAAACATTAACGGCGACTTTGCCTTGTTATTTAAT
CGCACTTGAAGGTAAAGGCGTTCACGTGGTAACCGTGAATGATTATCTTGCTCGCCGAGATGCAGAAACA
AACCGTCCGTTATTTGAATTTTTAGGCATGAGTGTAGGCGTCAATATTCCTGGTTTATCGCCAGAAGAAA
AACGTGCAGCTTATGCGGCAGATATTACTTATGCAACCAATAGTGAACTTGGTTTTGATTATTTACGTGA
CAACTTAGCCCACTCAAAAGAAGAGCGTTTCCAACGTACTTTAGGCTATGCGTTGGTGGATGAAGTGGAT
TCTATCTTAATCGATGAAGCGCGTACGCCATTGATTATTTCTGGTCAGGCAGAAAACAGTTCAGAGCTTT
ATATTGCGGTAAATAAATTGATCCCAAGTTTAATTAAACAAGAAAAAGAAGATACGGAAGAATATCAAGG
AGAGGGCGATTTCACTTTAGATTTGAAATCTAAACAAGCGCATTTAACCGAACGTGGTCAAGAAAAAGTA
GAAGATTGGTTAATTGCACAAGGTTTAATGCCTGAGGGGGACTCTTTGTATTCTCCTAGTCGAATTGTAT
TGCTTCATCACGTTATGGCTGCATTGCGTGCGCACACATTGTTTGAAAAAGATGTCGATTACATTGTGAA
GGACGGTGAAATCGTGATTGTTGATGAACACACTGGTCGTACAATGGCGGGGCGTCGTTGGTCAGATGGT
TTGCACCAAGCCATTGAGGCAAAAGAAGGGGTGGATGTTAAGAGCGAAAACCAAACTGTTGCATCAATTT
CTTACCAAAACTACTTCCGTTTATATGAACGTCTTGCGGGTATGACGGGGACTGCGGATACCGAAGCATT
TGAGTTCCAACAAATTTATGGCTTGGAAACTGTTGTAATTCCAACAAATCGTCCAATGATTCGTGATGAT
CGCACTGATGTGATGTTTGAAAATGAACAATATAAATTTAATGCGATTATTGAAGACATTAAAGATTGTG
TAGAACGCCAGCAACCAGTATTAGTGGGGACGATTTCAGTCGAAAAATCAGAAGAATTATCTAAAGCGTT
AGATAAAGCAGGTATAAAACACAATGTGTTGAATGCGAAATTCCACCAACAAGAAGCGGAAATCGTGGCA
GAAGCAGGATTTCCTAGCGCAGTGACTATCGCAACGAATATGGCGGGTCGAGGTACGGATATTATTCTTG
GCGGTAACTGGAAAGCGCAGGCTGCCAAATTAGAAAATCCAACTCAAGAACAAATTGAAGCCCTTAAAGC

SEQ ID NO:21

Fig. 5J (SHEET 1 OF 2)

AGAGTGGGAGAAAACCACGAAATTGTAATGAAAGCGGGTGGGTTGCATATTATCGGTACAGAGCGTCAC

GAATCTCGCCGTATTGATAACCAGTTGCGCGGTCGTTCTGGGCGTCAAGGTGACCCCGGTTCTTCTCGTT

TCTATCTTTCTTTGGAAGATGGTTTAATGCGCATTTATTTGAATGAGGGTAAGCTCAATTTAATGCGTAA

AGCGTTCACGGTAGCAGGCGAGGCAATGGAGTCGAAAATGTTGGCGAAAGTGATTGCATCTGCTCAAGCA

AAAGTTGAGGCGTTCCATTTTGATGGCCGTAAAAACCTACTTGAATATGATGATGTGGCAAATGACCAAC

GTCACGCGATTTATGAGCAACGCAATCATTTGCTTGATAATGATGATATTTCTGAAACTATCAACGCTAT

TCGCCACGATGTGTTTAATGGTGTGATTGATCAATATATTCCACCACAATCTTTGGAAGAACAATGGGAT

ATTAAAGGGCTTGAAGAACGTTTATCTCAAGAGTTTGGTATGGAATTACCGATTTCTAATTGGTTGGAAG

AAGATAATAATCTTCACGAAGAAAGTTTGCGCGAACGCATTGTGTGGAAATTGCAGAAAAGGAATACAAAGA

AAAAGAGGCTTTGGTTGGCGAAGACGCTATTGCGCCATTTTTGAAAAAGGTGTTATGTTGCAAACCTTAGAT

GAACTTTGGAAAGAACACTTAGCTTCGATGGATTATTTACGCCAAGGTATTCATTTACGTGGCTATGCCC

AAAAAGATCCAAAACAAGAGTATAAAAAAGAATCTTTCCGTATGTTTACGGAAATGTTGGATTCTTTAAA

ACACCAGGTTATCACGGCTTTAACCCGTGTACGTGCGTACTCAAGAAGAAATGGAAGAGCTGAACGT

GCTCGTCAAGAAATGGCAGCACGTATCAATCAAAATAATTTACCTGTGGATGAAAATAGTCAGACAACTC

AAAATTCAGAGACTGAAGATTATTCAGATCGTCGCATTGGTCGCAACGAGCCTTGTCCTTGTGGGTCGGG

TAAAAAATATAAGCATTGTCACGGCAGTCGTGTGGCACGCCAGTAA

SEQ ID NO:21

Fig. 5J (SHEET 2 OF 2)

>HI1650
ATGCCAAACGAACGTAATATTCAAAATTATCACTCGACTTACAACAACATTCGGGATTGGCTTGGTTATC
AAAAAGCTGGCGAGGAAAAAGCAAAGTCGACCATCAATTAG

SEQ ID NO:22

Fig. 5K

>HI1651
ATGGATGGCATATTACGTAAACTCATTTCAATTAAGGATTTACACCATTGCCTGCAGAAATTTTTTGTGG
ATGAAAGAGAATCTATTATAGAAATGAATGATAATAAGCTTTCAGAACAGTTTGATTTAGCATTGATTGA
AACGCATGGTAAATCAAAAATTTTAAAAAATTTATCTTTATTCAAACAAACCATGTCTAATTATCTTACT
CAATTATCAAAAGATAATATGAAAGAAACAGAAAATACTGTTCATAAAATTAAAAGAGTAGCAGCATAG

SEQ ID NO:23

Fig. 5L

>HI1654

ATGACTGATTTAACCGGAATTTTATACATTGTTGCCACGCCCATTGGCAATTTACAAGATATTACCCAAC

GTGCTTTAGAGACTTTTGCTCAAGTGGATTTAATTGCAGCAGAAGATACTCGCCATAGTGGACTTTTATT

GAGCCATTACGGCATTAAGAAGCCATTTTTTGCTTTGCACGATCATAACGAACACAAGAAAAAGCGCATATT

TTGGTGGAAAAGCTCAAGCAGGGAGTAATATTGCCTTTGATTTCTGATGCGGGGACGCCATTAATCAGTG

ACCCTGGTTTTCATTTAGTACGCCAATGCCGTGAAGCTGGCATTCGAGTTGTGCCTTTGCCAGGAGCTTG

TGCGGCAATTACCGCTCTTTGTGCATCGGGGATTTGCTTCTGATAGATTTTGTTTTGAAGGCTTTTTACCT

GCGAAAAGTAAAGCACGCAAAGATAAATTAGAAAATATCGCAGAAGAAGACCGCACTTTGATTTTTTATG

AATCCACTCACCGTATTTTAGATACACTAGAAGATATGCAAGCGGTGCTAGGGGAAGAACGATACATTGT

GTTAGCCCGTGAAATGACTAAAACTTGGGAAACGATTACGGGGAATACGATTAAAAATTTACGCGAATGG

CTTTTAGAAGATCCCAATCGTACAAAAGGCGAGATGGTTTTGATTGTGGAAGGCAAACCAAAGTCTGACA

ATAACGATGAAATTTCGCCGCAAGCGGTAAAGGCACTTGAGTTAATTGCAGAAGAATTGCCACTTAAAAA

AGCAGCAGCTATAGTTGCTGAGTTGTTATAAGAAGAATGCTTTGTATCAATTTGGATTAGCGCAT

TTGGAAAAATAA

SEQ ID NO:24

Fig. 5M

>HI1655
ATGTCTATTCTATTACAAGGCGAACGTTTTAAAAAACGTTTAATGCCAATTTTATTGTCAATGGCTTTAG
CTGGCTGTTCAAATCTACTTGGTAGCAATTTCACGCAAACCTTACAAAAAGATGCAAATGCAAGTTCTGA
ATTTTATATAAACAAATTAGGGCAAACACAAGAACTTGAAGATCAACAAACCTATAAATTGCTCGCGGCT
CGAGTGTTAATCCGTGAAAATAAGGTTGAACAATCGGCAGCGTTATTGAGGGAATTAGGCGAATTAAATG
ATGCGCAAAAATTAGATCGTGCATTAATTGAAGCGAGAATTTCTGCTGCAAAAAATGCCAATGAAGTCGC
ACAAAATCAATTACGTGCATTGGATTTAAATAAACTAAGCCCGTCACAAAAATCTCGTTATTACGAAACC
TTAGCTATTGTTGCCGAAAACCGTAAAGACATGATTGAAGCGGTAAAAGCGCGGATAGAAATGGATAAGA
ATTTAACAGATGTACAACGTCATCAAGATAATATTGATAAAACTTGGGCTTTATTGCGTTCAGCGAATAC
TGGCGTTATTAATAATGCCTCTGATGAAGGTAATGCAGCTTTAGGCGGTTGGCTAACATTAATCAAAGCC
TACAACGATTATATTCGTCAGCCTGTACAATTAAGCCAAGCCTTACAAAGTTGGAAAAATGCTTATCCAA
ATCATGCAGCCGCAACGTTGTTCCCAAAAGAATTGCTTACATTGCTTAATTTCCAACAAACGAATGTGTC
ACAAATTGGTTTACTCTTGCCATTAAGTGGTGACGGACAAATTCTTGGCACAACCATTCAATCGGGTTTT
AACGACGCGAAAGGTAACTCAACCATTCCAGTGCAAGTGTTTGATACCTCAATGAATTCTGTCCAAGATA
TCATTGCGCAAGCAAAACAAGCGGGGATTAAAACCTTAGTTGGCCCATTACTAAAACAAAATCTTGATGT
GATTTTAGCAGATCCTGCTCAAATTCAAGGTATGGATGTGCTTGCATTAAATGCCACACCAAATTCTCGT
GCGATTCCTCAACTTTGTTATTACGGACTTTCGCCAGAAGATGAAGCTGAATCTGCCGCCAATAAAATGT
GGAACGATGGCGTGCGTAATCCACTTGTCGCAATGCCGCAAAATGATTTAGGACAACGCGTAGGCAATGC
CTTTAATGTACGTTGGCAACAATTAGCAGGTACTGATGCGAATATCCGTTACTACAATTTGCCTGCGGAT
GTGACCTATTTCGTTCAAGAAAATAACTCAAATACAACCGCACTTTATGCCGTAGCAAGTCCAACTGAAC
TGGCAGAAATGAAAGGTTATTTAACAAATATCGTACCTAATTTAGCGATTTATGCCAGTTCTCGAGCAAG
CGCAAGTGCGACAAACACTAATACCGACTTCATCGCACAGATGAACGGTGTACAGTTTAGTGATATTCCA
TTTTTTAAAGATACCAATTCTCCACAATATCAGAAGTTAGCAAAATCCACGGGGGGCGAATATCAATTGA
TGCGTTTATATGCAATGGGTGCGGATGCGTGGTTGCTCATTAATCAATTTAATGAATTACGCCAAGTGCC
AGGCTATCGCTTGAGTGGCTTAACAGGGATTTTAAGTGCTGATACCAACTGTAATGTTGAACGCGATATG
ACTTGGTATCAATATCAAGATGGTGCAATTGTACCAGTTGCAAACTAA

SEQ ID NO:25

Fig. 5N

TCATGGCAAATCATTTCCGTACAGATCGTGTGGGCATGGAAATCAAGCGT
GAAGTCAATGAGATTTTGCAAAAGAAAGTCCGTGATCCACGTGTCCAAGG
TGTGACCATCATAGATGTTCAGATGCTGGGTGACTTGTCTGTTGCCAAGG
TTTATTACACCATTTTGAGTAACCTTGCTTCGGATAACCAAAAAGCCCAA
ATCGGGCTTGAAAAAGCAACTGGTACCATCAAACGTGAACTTGGTCGCAA
TTTGAAATTGTACAAAATCCCAGATTTGACCTTCGTCAAAGACGAGTCCA
TCGAGTATGGAAACAAGATTGACGAGATGCTACGCAATCTGGATAAG

rbfA (SEQ ID NO: 26)

Fig. 5O

GGATGTGAAAAGCCACTCATCAAGTGTGGAAGAAGCTGTCGCTGCAAAAATTGCTGCCAGCTTTAAGCCT
GCAGCTGCTCCGAAAGTAGAAGCAAAACCTGCAGCCCCAAAAGTAAGTGCAGAAAAGAAAGCCGAAAAAT
CTGAGCCAGCTAAACCAGCTGTAGCTAAGGAAGAGGCAAAACCTGCAGCCCCAAAAGCAAGTGCAGAAAA
GAAAGCCGAAAAGTCTGAACCAGTAAAACCAGCTGTAGCCAAGGAAGAGGCAAAACCAGCTGAGCCAGTC
ACTCCGAAAACAGAAAAAGTAGCGGCTAAACCGCAAAGTCGTAATTTCAAGGCTGAGCGTGAAGCACGTG
CTAAAGAGCAGGCAGAGCGACGCAAGCAAAATAAGGGCAATAACCGTGACCAACAACAAAACGGAAACCG
TCAGAAAAACGACGGCCGTAATGGTGGAAAACAAGGTCAAAGCAACCGCGACAATCGTCGCTTTAATGAC
CAAGCTAAGAAGCAGCAAGGTCAGCAAAAACGTAGAAATGAGCGCCGTCAGCAAGAGGATAAACGTTCAA
ATCAAGCGGCTCCACGTATTGACTTTAAAGCCCGTGCAGCAGCCCTAAAAGCAGAGCAAAATGCAGAGTA
CGCTCGTTCAAGTGAGGAACGCTTCAAGCAGTATCAGGCTGCTAAAGAAGCCTTGGCTCAAGCTAACAAA
CGCAAGGAACCAGAGGAAATCTTTGAAGAAGCGGCTAAGTTAGCTGAACAAGCACAGCAAGTTCAAGCAG
TGGTTGAAGTCGTCCCTGAGAAAAAAGAACCTGCAGTGGATACACGTCGTAAAAAACAAGCTCGACCAGA
CAAAAATCGTGACGATTATGATCATGAAGAAGATGGTCCTAGAAAACAACAAAAGAATCGAAGTAGTCAA
AATCAAGTGAGAAATCAAAAGAATAGTAACTGGAATAACAACAAAAAGAACAAAAAAGGCAATAACAAGA
ACAACCGTAATCAGACTCCAAAACCTGTTACGGAGCGTAAATTCCATGAATTGCCAACAGAATTTGAATA
TACAGATGGTATGACCGTTGCGGAAATCGCAAAACGTATCAAACGTGAACCAGCTGAAATTGTTAAGAAA
CTTTTCATGATGGGTGTCATGGCCACACAAAACCAATCCTTGGATGGGGAAACAATTGAACTCCTCATGG
TGGATTACGGTATCGAAGCCAAACAAAAGGTTGAAGTGGATAATGCTGACATCGAACGTTTCTTTGTCGA
AGATGGTTATCTCAATGAAGATGAATTGGTTGAGCGTCCACCAGTTGTTACTATCATGGGACACGTTGAC
CACGGTAAAACAACCCTTTTGGATACTCTTCGTAACTCACGTGTTGCGACAGGTGAAGCAGGTGGTATTA
CTCAGCATATCGGTGCCTACCAAATCGTGGAAAATGGTAAGAAGATTACCTTCCTTGATACACCAGGACA
CGCGGCCTTTACATCAATGCGTGCGCGTGGTGCTTCTGTTACCGATATTACGATCTTGGTCGTAGCGGCA
GATGACGGGGTTATGCCTCAGACTATTGAAGCCATCAACCACTCAAAAGCAGCTAACGTTCCAATCATCG
TAGCTATTAACAAGATTGATAAACCAGGTGCTAACCCAGAACGCGTTATCGGTGAATTGGCAGAGCATGG

IF-2 (SEQ ID NO: 27)

Fig. 5P (SHEET 1 OF 2)

TGTGATGTCAACTGCTTGGGGTGGAGATTCTGAATTTGTTGAAATTTCGGCTAAATTCAACCAAAATATC
GAAGAATTGTTGGAAACAGTCCTTCTTGTGGCTGAAATCCAAGAACTCAAAGCAGACCCAACAGTTCGTG
CGATCGGTACGGTTATCGAAGCGCGCTTGGATAAAGGAAAAGGTGCGGTCGCAACCCTTCTTGTACAACA
AGGTACCTTGAATGTTCAAGACCCAATCGTTGTCGGAAATAcTTCGGTCGTGTCCGTGCTATGACCAACG
ACCTTGGTCGTCGTGTTAAAGTTGCTGGACCATCAACACCAGTCTCTATCACAGGTTTGAACGAAGCACC
GATGGCGGGTGACCACTTTGCCGTTTACGAGGATGAAAAATCTGCGCGTGCAGCAGGTGAAGAGCGTGCC
AAACGTGCCCTCATGAAACAACGTCAAGCTACCCAACGTGTTAGCCTTGAAAACCTCTTTGATACCCTTA
AAGCTGGGGAACTCAAATCTGTTAATGTTATCATCAAGGCTGATGTACAAGGTTCTGTTGAAGCCCTTTC
TGCCTCACTTCAAAAGATTGACGTGGAAGGTGTCAAAGTGACTATCGTCCACTCAGCGGTCGGTGCTATC
AACGAATCAGACGTGACCCTTGCCGAAGCTTCAAATGCCTTTATCGTTGGTTTCAACGTACGCCCTACAC
CACAAGCTCGTCAACAAGCAGAAGCTGACGATGTGGAAATCCGTCTTCACAGCATTATCTACAAGGTTAT
CGAAGAGATGGAAGAAGCTATGAAAGGGATGCTTGATCCAGAATTTGAAGAAAAAGTTATTGGTGAAGCG
GTTATCCGTGAAACCTTCAAGGTGTCTAAAGTGGGAACTATCGGTGGATTTATGGTTATCAACGGTAAGG
TTGCCCGTGACTCTAAAGTCCGTGTTATCCGTGATGGTGTCGTTATCTATGATGGTGAACTCGCAAGCTT
GAAACACTATAAAGACGACGTGAAAGAAGTGACAAACGGTCGTGAAGGTGGATTGATGATCGACGGCTAC
AATGATATTAAGATGGATGATGTGATTGAGGCGTATGTCATGGAAGAAATCAAGAGA

IF-2 (SEQ ID NO: 27)

Fig. 5P (SHEET 2 OF 2)

AATAAGCAAAAGATAAGTAATCTCTTGGGGCTTGCTCAGCGAGCAGGGCG
CATCATATCGGGTGAAGAATTGGTGGTCAAGGCCATTCAAGACGGCAAGG
CCAAGTTGGTCTTTCTAGCTCATGATGCTGGACCCAATCTGACCAAGAAG
ATTCAAGATAAAAGTCATTATTATCAAGTAGAAATTGTAACCGTGTTTTC
AACACTGGAATTAAGCATAGCAGTCGGGAAATCGAGAAAGGTTTTTGGCTG
TAACAGATGCTGGATTTACAAAGAAAATGAGGTCTCTTATGGAA

L7AE (SEQ ID NO: 28)

Fig. 5Q

ATGAGTAAAGAAATGCTAGAGGCCTTCCGCATTTTGGAAGAAGACAAGGG
AATCAAAAAAGAAGATATCATCGACGCAGTAGTAGAGTCGCTTCGTTCCG
CTTATCGCAGACGGCTATGGTCAGTCAGACAGCGTAGCTATTGACTTCAAC
GAAAAAACAGGTGACTTTACAGTTTATACTGTCCGTGAAGTTGTTGATGA
AGTATTTGATAGCCGTTTGGAAATCAGCTTGAAAGATGCTCTTGCCATTA
ATTCAGCTTATGAACTTGGAGACAAAAATCAAGTTTGAAGAAGCACCAGCT
GAGTTTGGTCGTGTAGCAGCCCAATCTGCCAAACAAACCATCATGGAAAA
AATGCGCAaCAAACACGTGCCATCACTTACAATACTTACAAAGAACATGA
GCAAGAAATCATGTCTGGTACAGTAGAACGCTTTGACAACCGCTTTATCT
ATGTCAACCTTGGTAGCATCGAAGCCCAATTGTCAAAACAAGACCAAATT
CCTGGAGAAGTTTTTGCTTCTCATGATCGTATCGAAGTTTATGTTTACAA
GGTTGAAGACAACCCTCGTGGTGTGAACGTCTTTGTTAGCCGTAGTCATC
CAGAAATGATCAAACGTTTAATGGAGCAAGAAATTCCAGAAGTTTATGAT
GGAACTGTTGAAATCATGAGCGTGGCTCGTGAAGCAGGTGACCGTACGAA
GGTTGCTGTTCGTAGCCACAATCCAAACGTGGATGCTATCGGTACAATCG
TTGGACGTGGTGGTGCTAATATCAAGAAGATTACTAGCAAATTCCACCCA
GCTCGTTACGATGCTAAAAATGACCGCATGGTACCAATCGAAGAAAATAT
CGATGTTATCGAGTGGGTAGCAGATCCAGCTGAATTTATCTACAATGCCA
TCGCTCCTGCTGAGGTTGACCAAGTTATCTTTGATGAAAACGACAGCAAA
CGTGCCTTGGTGGTTGTTCCAGATAACAAGCTTTCTCTTGCCATTGGTCG
TCGTGGACAAAACGTGCGCTTGGCGGCTCACTTGACTGGTTACCGTATCG
ATATCAAGTCTGCTAGCGAA

nusA (SEQ ID NO: 29)

Fig. 5R

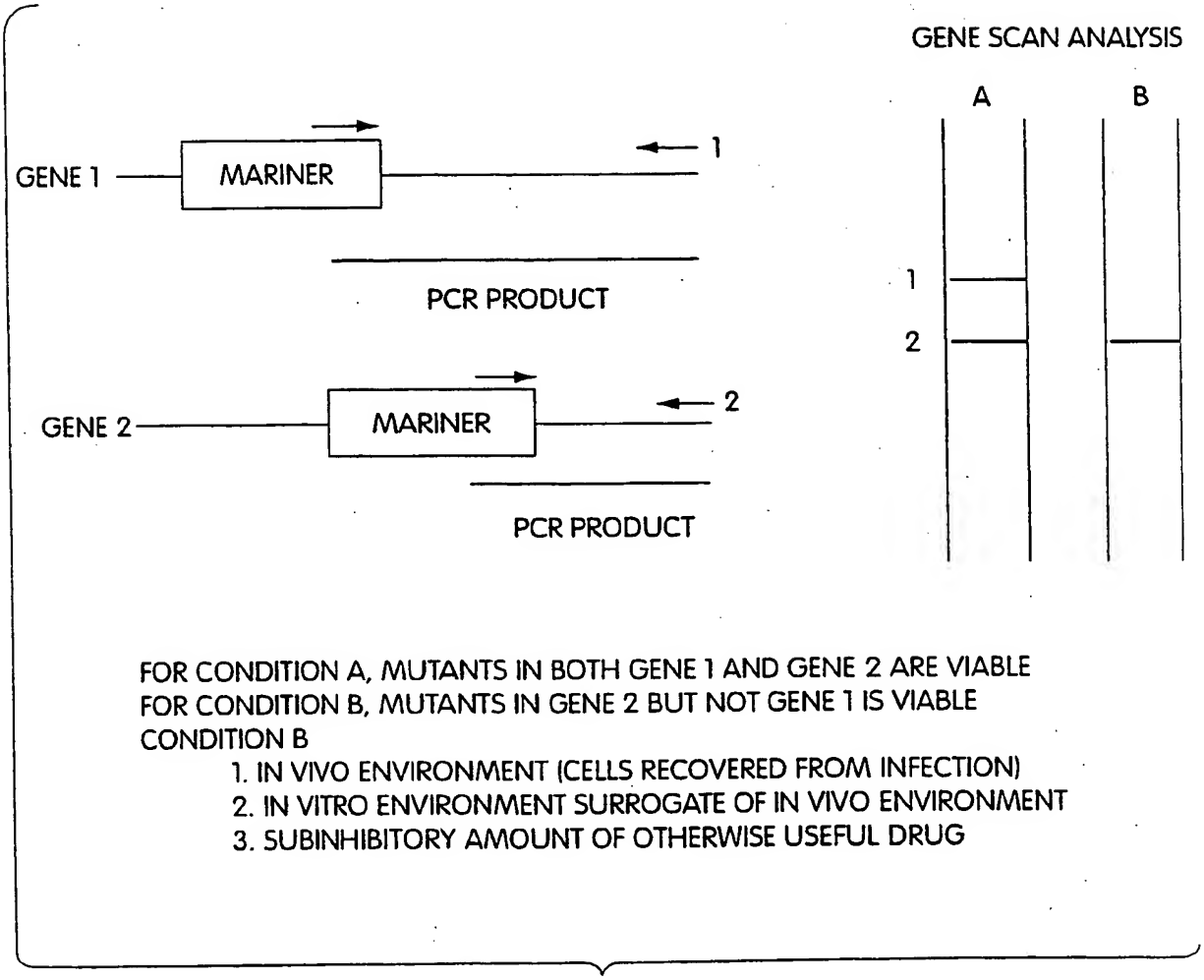
GENE SCAN ANALYSIS
A
B
GENE 1
MARINER
1
PCR PRODUCT
1
2
GENE 2
MARINER
2
PCR PRODUCT
FOR CONDITION A, MUTANTS IN BOTH GENE 1 AND GENE 2 ARE VIABLE
FOR CONDITION B, MUTANTS IN GENE 2 BUT NOT GENE 1 IS VIABLE
CONDITION B
1. IN VIVO ENVIRONMENT (CELLS RECOVERED FROM INFECTION)
2. IN VITRO ENVIRONMENT SURROGATE OF IN VIVO ENVIRONMENT
3. SUBINHIBITORY AMOUNT OF OTHERWISE USEFUL DRUG

Fig. 6

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US99/06139

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) : C12N 15/00
US CL : 435/172.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/172.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

STN (Medline, EUROPATFULL, Biosis, CAPLUS, Lifesci, Embase) and US PATENTS (APS)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	AKERLY, B.J. et al. Systematic Identification of Essential Genes by In Vitro Mariner Mutagenesis. Proc. Natl. Acad. Sci, USA. July 1998, entire document.	1-24
A	KURTZ, S. et al. Growth Impairment Resulting from Expression of Influenza Virus M2 Protein in Saccharomyces cerevisae: Identification of a Novel Inhibitor of Influenza Virus. Antimicrobial agents and Chemotheray. October 1995, Vol. 39, pages 2204-2209, entire document.	1-24
Y	US 5,173,294 A (MURPHY et al.) 22 December 1992, especially Abstract.	5, 6, 17, 18, 19, 20-24

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
07 JUNE 1999

Date of mailing of the international search report
08 JUL 1999

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
JANELL TAYLOR
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/06139

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P --- Y,P	US 5,843,772 A (DEVINE et al) 01 December 1998, Columns 10 and 11.	1-4, 10-15, 20-22 --- 5-8, 16-19, 23-24
X,P --- Y,P	US 5,792,633 A (SCHIESTL et al) 11 August 1998, Summary, Col. 6.	1-4, 10-15, 20-22 --- 5-8, 16-19, 23-24
X,P --- Y,P	US 5,817,502 A (LIGON et al) 06 October 1998, entire document, especially Col. 14.	1-4, 10-18, 20-22 --- 5-8, 19, 23-24